2. Random Variables

Dave Goldsman

Georgia Institute of Technology, Atlanta, GA, USA

5/21/14

Outline

1. Intro / Definitions
2. Discrete Random Variables
3. Continuous Random Variables
4. Cumulative Distribution Functions
5. Great Expectations
6. Functions of a Random Variable
7. Bivariate Random Variables
8. Conditional Distributions
9. Independent Random Variables
10. Conditional Expectation
11. Covariance and Correlation
12. Moment Generating Functions

Definition: A **random variable** (RV) is a function from the sample space to the real line. $X : S \to \mathbb{R}$.

Example: Flip 2 coins. $S = \{HH, HT, TH, TT\}$.

Suppose X is the RV corresponding to the # of H's.

$X(TT) = 0$, $X(HT) = X(TH) = 1$, $X(HH) = 2$.

$P(X = 0) = \frac{1}{4}$, $P(X = 1) = \frac{1}{2}$, $P(X = 2) = \frac{1}{4}$.

Notation: Capital letters like X, Y, Z, U, V, W usually represent RV's.

Small letters like x, y, z, u, v, w usually represent particular values of the RV's.

Thus, you can speak of $P(X = x)$.

Example: Let X be the sum of two dice rolls. Then $X((4,6)) = 10$. In addition,

$$P(X = x) \;=\; \begin{cases} 1/36 & \text{if } x = 2 \\ 2/36 & \text{if } x = 3 \\ \quad\vdots & \\ 6/36 & \text{if } x = 7 \\ \quad\vdots & \\ 1/36 & \text{if } x = 12 \\ \;\;0 & \text{otherwise} \end{cases}$$

Example: Flip a coin.

$$X \equiv \begin{cases} 0 & \text{if } T \\ 1 & \text{if } H \end{cases}$$

Example: Roll a die.

$$Y \equiv \begin{cases} 0 & \text{if } \{1,2,3\} \\ 1 & \text{if } \{4,5,6\} \end{cases}$$

For our purposes, X and Y are the same, since $P(X=0) = P(Y=0) = \frac{1}{2}$ and $P(X=1) = P(Y=1) = \frac{1}{2}$.

Example: Select a real # at random betw 0 and 1.

Infinite number of “equally likely” outcomes.

Conclusion: $P($each *individual point*$) = P(X = x) = 0$, believe it or not!

But $P(X \leq 0.5) = 0.5$ and $P(X \in [0.3, 0.7]) = 0.4$.

If A is any *interval* in [0,1], then $P(A)$ is the length of A.

Definition: If the number of possible values of a RV X is finite or countably infinite, then X is a **discrete** RV. Otherwise,...

A **continuous** RV is one with prob 0 at every point.

Example: Flip a coin — get H or T. Discrete.

Example: Pick a point at random in $[0, 1]$. Continuous.

Outline

1. Intro / Definitions
2. Discrete Random Variables
3. Continuous Random Variables
4. Cumulative Distribution Functions
5. Great Expectations
6. Functions of a Random Variable
7. Bivariate Random Variables
8. Conditional Distributions
9. Independent Random Variables
10. Conditional Expectation
11. Covariance and Correlation
12. Moment Generating Functions

Definition: If X is a discrete RV, its **probability mass function** (pmf) is $f(x) \equiv P(X = x)$.

Note that $0 \leq f(x) \leq 1$, $\sum_x f(x) = 1$.

Example: Flip 2 coins. Let X be the number of heads.

$$f(x) = \begin{cases} 1/4 & \text{if } x = 0 \text{ or } 2 \\ 1/2 & \text{if } x = 1 \\ 0 & \text{otherwise} \end{cases}$$

Example: **Uniform Distribution** on integers $1, 2, \ldots, n$. X can equal $1, 2, \ldots, n$, each with prob $1/n$. $f(i) = 1/n$, $i = 1, 2, \ldots, n$.

Example/Definition: Let X denote the number of "successes" from n independent trials such that the $P(\text{success})$ at each trial is p $(0 \leq p \leq 1)$. Then X has the **Binomial Distribution** with parameters n and p.

The trials are referred to as **Bernoulli trials.**

Notation: $X \sim \text{Bin}(n, p)$. "X has the Bin distribution"

Example: Roll a die 3 indep times. Find

$$P(\text{Get exactly two 6's}).$$

"success" (6) and "failure" (1,2,3,4,5)

All 3 trials are indep, and $P(\text{success}) = 1/6$ doesn't change from trial to trial.

Let $X =$ # of 6's. Then $X \sim \text{Bin}(3, \frac{1}{6})$.

Theorem: If $X \sim \text{Bin}(n, p)$, then the prob of k successes in n trials is

$$P(X = k) \;=\; \binom{n}{k} p^k q^{n-k}, \quad k = 0, 1, \ldots, n,$$

where $q = 1 - p$.

Proof: Particular sequence of successes and failures:

$$\underbrace{SS \cdots S}_{k \text{ successes}} \; \underbrace{FF \cdots F}_{n-k \text{ failures}} \quad (\text{prob} = p^k q^{n-k})$$

Number of ways to arrange the seq is $\binom{n}{k}$. Done.

Back to the dice example, where $X \sim \mathrm{Bin}(3, \frac{1}{6})$, and we want $P(\text{Get exactly two 6's})$.

$n = 3$, $k = 2$, $p = 1/6$, $q = 5/6$.

$$P(X = 2) \;=\; \binom{3}{2}\left(\frac{1}{6}\right)^2\left(\frac{5}{6}\right)^1 \;=\; \frac{15}{216}$$

k	0	1	2	3
$P(X = k)$	$\frac{125}{216}$	$\frac{75}{216}$	$\frac{15}{216}$	$\frac{1}{216}$

Example: Roll 2 dice 12 times.

Find P(Result will be 7 or 11 exactly 3 times).

Let $X =$ the number of times get 7 or 11.

$P(\text{7 or 11}) \;=\; P(7) + P(11) \;=\; \frac{6}{36} + \frac{2}{36} \;=\; \frac{2}{9}$.

So $X \sim \text{Bin}(12, 2/9)$.

$$P(X = 3) \;=\; \binom{12}{3}\left(\frac{2}{9}\right)^3\left(\frac{7}{9}\right)^9.$$

Definition: If $P(X = k) = e^{-\lambda}\lambda^k/k!$, $k = 0, 1, 2, \ldots$, $\lambda > 0$, we say that X has the **Poisson distribution** with parameter λ.

Notation: $X \sim \text{Pois}(\lambda)$.

Example: Suppose the number of raisins in a cup of cookie dough is Pois(10). Find the prob that a cup of dough has at least 4 raisins.

$$\begin{aligned} P(X \geq 4) &= 1 - P(X = 0, 1, 2, 3) \\ &= 1 - e^{-10}\left(\frac{10^0}{0!} + \frac{10^1}{1!} + \frac{10^2}{2!} + \frac{10^3}{3!}\right) \\ &= 0.9897. \end{aligned}$$

Outline

1. Intro / Definitions
2. Discrete Random Variables
3. Continuous Random Variables
4. Cumulative Distribution Functions
5. Great Expectations
6. Functions of a Random Variable
7. Bivariate Random Variables
8. Conditional Distributions
9. Independent Random Variables
10. Conditional Expectation
11. Covariance and Correlation
12. Moment Generating Functions

Example: Pick a point X randomly between 0 and 1.

$$f(x) = \begin{cases} 1 & \text{if } 0 \leq x \leq 1 \\ 0 & \text{otherwise} \end{cases}$$

$$\begin{aligned} P(a < X < b) &= \text{area under } f(x) \text{ from } a \text{ to } b \\ &= b - a. \end{aligned}$$

Definition: Suppose X is a continuous RV. $f(x)$ is the **probability density function** (pdf) if

- $\int_{\mathbb{R}} f(x)\,dx = 1$ (area under $f(x)$ is 1)
- $f(x) \geq 0, \quad \forall x$ (always non-negative)
- If $A \subseteq \mathbb{R}$, then $P(X \in A) = \int_A f(x)\,dx$ (probability that X is in a certain region A)

Remarks: If X is a continuous RV, then

$P(a < X < b) = \int_a^b f(x)\,dx$.

An individual point has prob 0, i.e., $P(X = x) = 0$.

Think of $f(x)\,dx \approx P(x < X < x + dx)$.

Note that $f(x)$ denotes both pmf (**discrete** case) and pdf (**continuous** case) — but they are **different**:

$f(x) = P(X = x)$ if X is **discrete.**
Must have $0 \leq f(x) \leq 1$.

$f(x)\,dx \approx P(x < X < x + dx)$ if X is **continuous.** Must have $f(x) \geq 0$ (and possibly > 1).

If X is cts, we calculate the prob of an event by integrating, $P(X \in A) = \int_A f(x)\,dx$.

Example: If X is "equally likely" to be anywhere between a and b, then X has the **uniform distribution** on (a, b).

$$f(x) = \begin{cases} \frac{1}{b-a} & \text{if } a < x < b \\ 0 & \text{otherwise} \end{cases}$$

Notation: $X \sim \mathrm{U}(a, b)$

Note: $\int_{\mathbb{R}} f(x)\,dx = \int_a^b \frac{1}{b-a}\,dx = 1$ (as desired).

Example: X has the **exponential distriibution** with parameter $\lambda > 0$ if it has pdf

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Notation: $X \sim \mathrm{Exp}(\lambda)$

Note: $\int_{\mathbb{R}} f(x)\,dx = \int_0^\infty \lambda e^{-\lambda x}\,dx = 1$ (as desired).

Example: Suppose $X \sim \text{Exp}(1)$. Then

$P(X \leq 3) = \int_0^3 e^{-x}\,dx = 1 - e^{-3}$.

$P(X \geq 5) = \int_5^\infty e^{-x}\,dx = e^{-5}$.

$P(2 \leq X < 4) = P(2 \leq X \leq 4) = \int_2^4 e^{-x}\,dx = e^{-2} - e^{-4}$.

$P(X = 3) = \int_3^3 e^{-x}\,dx = 0$.

Example: Suppose X is a cts RV with pdf

$$f(x) = \begin{cases} cx^2 & \text{if } 0 < x < 2 \\ 0 & \text{otherwise} \end{cases}$$

Find c.

Answer: $1 = \int_{\mathbb{R}} f(x)\,dx = \int_0^2 cx^2\,dx = \frac{8}{3}c$, so $c = 3/8$.

Find $P(0 < X < 1)$.

Answer: $P(0 < X < 1) = \int_0^1 \frac{3}{8}x^2\,dx = 1/8$.

Find $P(0 < X < 1|\frac{1}{2} < X < \frac{3}{2})$.

Answer:

$$P\left(0 < X < 1 \middle| \frac{1}{2} < X < \frac{3}{2}\right)$$

$$= \frac{P(0 < X < 1 \text{ and } \frac{1}{2} < X < \frac{3}{2})}{P(\frac{1}{2} < X < \frac{3}{2})}$$

$$= \frac{P(\frac{1}{2} < X < 1)}{P(\frac{1}{2} < X < \frac{3}{2})}$$

$$= \frac{\int_{1/2}^{1} \frac{3}{8}x^2\,dx}{\int_{1/2}^{3/2} \frac{3}{8}x^2\,dx} = 7/26.$$

Outline

1. Intro / Definitions
2. Discrete Random Variables
3. Continuous Random Variables
4. Cumulative Distribution Functions
5. Great Expectations
6. Functions of a Random Variable
7. Bivariate Random Variables
8. Conditional Distributions
9. Independent Random Variables
10. Conditional Expectation
11. Covariance and Correlation
12. Moment Generating Functions

Definition: For any RV X, the **cumulative distribution function** (cdf) is defined for all x by $F(x) \equiv P(X \leq x)$.

X discrete implies

$$F(x) \;=\; \sum_{\{y|y\leq x\}} f(y) \;=\; \sum_{\{y|y\leq x\}} P(X=y).$$

X continuous implies

$$F(x) \;=\; \int_{-\infty}^{x} f(t)\,dt.$$

Discrete cdf's

Example: Flip a coin twice. $X =$ number of H's.

$$X = \begin{cases} 0 \text{ or } 2 & \text{w.p. } 1/4 \\ 1 & \text{w.p. } 1/2 \end{cases}$$

$$F(x) = P(X \leq x) = \begin{cases} 0 & \text{if } x < 0 \\ 1/4 & \text{if } 0 \leq x < 1 \\ 3/4 & \text{if } 1 \leq x < 2 \\ 1 & \text{if } x \geq 2 \end{cases}$$

(You have to be careful about "$\leq$" vs. "$<$".)

Continuous cdf's

Theorem: If X is a **continuous** RV, then $f(x) = F'(x)$.

Proof: $F'(x) = \frac{d}{dx} \int_{-\infty}^{x} f(t)\,dt = f(x)$, by the fundamental theorem of calculus.

Remark: If X is continuous, you can get from the pdf $f(x)$ to the cdf $F(x)$ by integrating.

Example: $X \sim \mathrm{U}(0,1)$.

$$f(x) = \begin{cases} 1 & \text{if } 0 < x < 1 \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = \begin{cases} 0 & \text{if } x \le 0 \\ x & \text{if } 0 < x < 1 \\ 1 & \text{if } x \ge 1 \end{cases}$$

Example: $X \sim \text{Exp}(\lambda)$.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

$$F(x) = \int_{-\infty}^{x} f(t)\,dt = \begin{cases} 0 & \text{if } x \le 0 \\ 1 - e^{-\lambda x} & \text{if } x > 0 \end{cases}$$

Properties of all cdf's

$F(x)$ is *non-decreasing* in x, i.e., $a < b$ implies that $F(a) \leq F(b)$.

$\lim_{x\to\infty} F(x) = 1$ and $\lim_{x\to-\infty} F(x) = 0$.

$F(x)$ is *right-cts* at every point x.

Theorem: $P(X > x) = 1 - F(x)$.

Proof:

$$1 \; = \; P(X \leq x) + P(X > x) \; = \; F(x) + P(X > x).$$

Theorem:

$$a < b \;\Rightarrow\; P(a < X \le b) \;=\; F(b) - F(a).$$

Proof:

$$\begin{aligned} & P(a < X \le b) \\ &= P(X > a \,\cap\, X \le b) \\ &= P(X > a) + P(X \le b) - P(X > a \,\cup\, X \le b) \\ &= 1 - F(a) + F(b) - 1. \end{aligned}$$

Outline

1. Intro / Definitions
2. Discrete Random Variables
3. Continuous Random Variables
4. Cumulative Distribution Functions
5. Great Expectations
6. Functions of a Random Variable
7. Bivariate Random Variables
8. Conditional Distributions
9. Independent Random Variables
10. Conditional Expectation
11. Covariance and Correlation
12. Moment Generating Functions

Great Expectations

Mean (Expected Value)

Law of the Unconscious Statistician

Variance

Chebychev's Inequality

Definition: The **mean** or **expected value** or **average** of a RV X is

$$\mu \equiv \mathrm{E}[X] \equiv \begin{cases} \sum_x x f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} x f(x)\, dx & \text{if } X \text{ is cts} \end{cases}$$

The mean gives an indication of a RV's *central tendency*. It can be thought of as a weighted average of the possible x's, where the weights are given by $f(x)$.

Example: Suppose X has the **Bernoulli distribution** with parameter p, i.e., $P(X=1)=p$, $P(X=0)=q=1-p$. Then

$$\mathrm{E}[X] \;=\; \sum_x xP(X=x) \;=\; 1\cdot p + 0\cdot q \;=\; p.$$

Example: Die. $X = 1,2,\ldots,6$, each w.p. 1/6. Then

$$\mathrm{E}[X] \;=\; \sum_x xf(x) \;=\; 1\cdot\frac{1}{6}+\cdots+6\cdot\frac{1}{6} \;=\; 3.5.$$

Example: Suppose X has the **Geometric distribution** with parameter p, i.e., X is the number of Bern(p) trials until you obtain your first success (e.g., $FFFFS$ corresponds to $X=5$). Then $f(x)=(1-p)^{x-1}p$, $x=1,2,\ldots$, and after a *lot* of algebra,

$$\mathrm{E}[X] \;=\; \sum_x xP(X=x) \;=\; \sum_{x=1}^{\infty} x(1-p)^{x-1}p \;=\; 1/p.$$

Example: $X \sim \mathrm{Exp}(\lambda)$. $f(x) = \lambda e^{-\lambda x}$, $x \geq 0$. Then

$$\begin{aligned}
\mathrm{E}[X] &= \int_{\mathbb{R}} x f(x)\, dx \\
&= \int_0^\infty x \lambda e^{-\lambda x}\, dx \\
&= -x e^{-\lambda x}\big|_0^\infty - \int_0^\infty (-e^{-\lambda x})\, dx \quad \text{(by parts)} \\
&= \int_0^\infty e^{-\lambda x}\, dx \quad \text{(L'Hôpital's rule)} \\
&= 1/\lambda.
\end{aligned}$$

Law of the Unconscious Statistician (LOTUS)

Definition/Theorem: The expected value of a function of X, say $h(X)$, is

$$\mathrm{E}[h(X)] \equiv \begin{cases} \sum_x h(x)f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} h(x)f(x)\,dx & \text{if } X \text{ is cts} \end{cases}$$

$\mathrm{E}[h(X)]$ is simply a weighted function of $h(x)$, where the weights are the $f(x)$ values.

Examples: $\mathrm{E}[X^2] = \int_{\mathbb{R}} x^2 f(x)\,dx$

$\mathrm{E}[\sin X] = \int_{\mathbb{R}} (\sin x) f(x)\,dx$

Just a moment please. . .

Definition: The kth **moment** of X is

$$E[X^k] = \begin{cases} \sum_x x^k f(x) & \text{if } X \text{ is discrete} \\ \int_{\mathbb{R}} x^k f(x)\,dx & \text{if } X \text{ is cts} \end{cases}$$

Definition: The kth **central moment** of X is

$$E[(X-\mu)^k] = \begin{cases} \sum_x (x-\mu)^k f(x) & X \text{ is discrete} \\ \int_{\mathbb{R}} (x-\mu)^k f(x)\,dx & X \text{ is cts} \end{cases}$$

Definition: The **variance** of X is the second central moment, i.e., $\mathrm{Var}(X) \equiv \mathrm{E}[(X-\mu)^2]$. It's a measure of spread or dispersion.

Notation: $\sigma^2 \equiv \mathrm{Var}(X)$.

Definition: The **standard deviation** of X is $\sigma \equiv +\sqrt{\mathrm{Var}(X)}$.

Theorem: For any $h(X)$ and constants a and b, we have $\mathrm{E}[ah(X)+b] = a\mathrm{E}[h(X)]+b$.

Proof (just do cts case):

$$\begin{aligned}
\mathrm{E}[ah(X)+b] &= \int_{\mathbb{R}} (ah(x)+b)f(x)\,dx \\
&= a\int_{\mathbb{R}} h(x)f(x)\,dx + b\int_{\mathbb{R}} f(x)\,dx \\
&= a\mathrm{E}[h(X)]+b.
\end{aligned}$$

Remark: In particular, $\mathrm{E}[aX+b] = a\mathrm{E}[X]+b$.

Similarly, $\mathrm{E}[g(X)+h(X)] = \mathrm{E}[g(X)]+\mathrm{E}[h(X)]$.

Theorem (easier way to calculate variance):

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2$$

Proof:

$$\begin{aligned}
\mathrm{Var}(X) &= \mathrm{E}[(X-\mu)^2] \\
&= \mathrm{E}[X^2 - 2\mu X + \mu^2] \\
&= \mathrm{E}[X^2] - 2\mu\mathrm{E}[X] + \mu^2 \quad \text{(by above remarks)} \\
&= \mathrm{E}[X^2] - \mu^2.
\end{aligned}$$

Example: $X \sim \text{Bern}(p)$.

$$X = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } q \end{cases}$$

Recall $\text{E}[X] = p$. In fact, for any k,

$$\text{E}[X^k] = 0^k \cdot q + 1^k \cdot p = p.$$

So $\text{Var}(X) = \text{E}[X^2] - (\text{E}[X])^2 = p - p^2 = pq$.

Example: $X \sim \mathrm{U}(a,b)$. $f(x) = \frac{1}{b-a}$, $a < x < b$.

$$\mathrm{E}[X] = \int_a^b x \frac{1}{b-a}\, dx = \frac{a+b}{2}$$

$$\mathrm{E}[X^2] = \int_a^b x^2 \frac{1}{b-a}\, dx = \frac{a^2+ab+b^2}{3}$$

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \frac{(a-b)^2}{12} \text{ (algebra).}$$

Theorem: $\mathrm{Var}(aX+b) = a^2\mathrm{Var}(X)$. A "shift" doesn't matter!

Proof:

$$\begin{aligned}
\mathrm{Var}(aX+b) &= \mathrm{E}[(aX+b)^2] - (\mathrm{E}[aX+b])^2 \\
&= \mathrm{E}[a^2X^2 + 2abX + b^2] - (a\mathrm{E}[X]+b)^2 \\
&= a^2\mathrm{E}[X^2] + 2ab\mathrm{E}[X] + b^2 \\
&\qquad\qquad -(a^2(\mathrm{E}[X])^2 + 2ab\mathrm{E}[X] + b^2) \\
&= a^2(\mathrm{E}[X^2] - (\mathrm{E}[X])^2) \\
&= a^2\mathrm{Var}(X)
\end{aligned}$$

Example: $X \sim \text{Bern}(0.3)$

Recall that $\text{E}[X] = p = 0.3$ and
$\text{Var}(X) = pq = (0.3)(0.7) = 0.21$.

Let $Y = h(X) = 4X + 5$. Then

$\text{E}[Y] = \text{E}[4X + 5] = 4\text{E}[X] + 5 = 6.2$

$\text{Var}(Y) = \text{Var}(4X + 5) = 16\text{Var}(X) = 3.36$.

Approximations to $\mathrm{E}[h(X)]$ and $\mathrm{Var}(h(X))$

Sometimes $Y = h(X)$ is messy, and we may have to approximate $\mathrm{E}[h(X)]$ and $\mathrm{Var}(h(X))$. Use a Taylor series approach. To do so, let $\mu = \mathrm{E}[X]$ and $\sigma^2 = \mathrm{Var}(X)$ and note that

$$Y = h(\mu) + (X-\mu)h'(\mu) + \frac{(X-\mu)^2}{2}h''(\mu) + R,$$

where R is a remainder term which we shall now ignore.

Then

$$\mathrm{E}[Y] \doteq h(\mu) + E[X-\mu]h'(\mu) + \frac{\mathrm{E}[(X-\mu)^2]}{2}h''(\mu) = h(\mu) + \frac{h''(\mu)\sigma^2}{2}$$

and (now we'll use an even-cruder approximation)

$$\mathrm{Var}(Y) \doteq \mathrm{Var}\big(h(\mu) + (X-\mu)h'(\mu)\big) = [h'(\mu)]^2\sigma^2.$$

Approximations to $\mathrm{E}[h(X)]$ and $\mathrm{Var}(h(X))$

Example: Suppose that X has pdf $f(x) = 3x^2$, $0 \le x \le 1$, and that we want to test out out approximations on the “complicated” random variable $Y = h(X) = X^{3/4}$. Well, it’s not really that complicated, since we can calculate the exact moments:

$$\begin{aligned}
\mathrm{E}[Y] &= \int_0^1 x^{3/4} f(x)\,dx = \int_0^1 3x^{11/4}\,dx = 4/5 \\
\mathrm{E}[Y^2] &= \int_0^1 x^{6/4} f(x)\,dx = \int_0^1 3x^{7/2}\,dx = 2/3 \\
\mathrm{Var}(Y) &= \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 = 2/75 = 0.0267.
\end{aligned}$$

Before we can do the approximation, note that

$$\begin{aligned}
\mu = \mathrm{E}[X] &= \int_0^1 x f(x)\,dx = \int_0^1 3x^3\,dx = 3/4 \\
\sigma^2 = \mathrm{Var}(X) &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = 3/80 = 0.0375.
\end{aligned}$$

Further,

$$\begin{aligned} h(\mu) &= \mu^{3/4} = (3/4)^{3/4} = 0.8059 \\ h'(\mu) &= (3/4)\mu^{-1/4} = (3/4)(3/4)^{-1/4} = 0.8059 \\ h''(\mu) &= -(3/16)\mu^{-5/4} = -0.2686. \end{aligned}$$

Thus,

$$\mathrm{E}[Y] \doteq h(\mu) + \frac{h''(\mu)\sigma^2}{2} = 0.8059 - \frac{(0.2686)(0.0375)}{2} = 0.8009$$

and

$$\mathrm{Var}(Y) \doteq [h'(\mu)]^2\sigma^2 = (0.8059)^2(0.0375) = 0.0243,$$

both of which are reasonably close to their true values.

Chebychev's Inequality

Theorem: Suppose that $\mathrm{E}[X] = \mu$ and $\mathrm{Var}(X) = \sigma^2$. Then for any $\epsilon > 0$,

$$P(|X - \mu| \geq \epsilon) \ \leq \ \sigma^2/\epsilon^2.$$

Proof: See text.

Remarks: If $\epsilon = k\sigma$, then $P(|X - \mu| \geq k\sigma) \leq 1/k^2$.

$P(|X - \mu| < \epsilon) \geq 1 - \sigma^2/\epsilon^2$.

Chebychev gives a **crude** bound on the prob that X deviates from the mean by more than a constant, in terms of the constant and the variance. You can always use Chebychev, but it's crude.

Example: Suppose $X \sim \mathrm{U}(0,1)$. $f(x) = 1$, $0 < x < 1$.

$\mathrm{E}[X] = \frac{a+b}{2} = 1/2$, $\mathrm{Var}(X) = \frac{(a-b)^2}{12} = 1/12$.

Then Chebychev implies

$$P\left(\left|X - \frac{1}{2}\right| \geq \epsilon\right) \ \leq \ \frac{1}{12\epsilon^2}.$$

In particular, for $\epsilon = 1/3$,

$$P\left(\left|X - \frac{1}{2}\right| \geq \frac{1}{3}\right) \ \leq \ \frac{3}{4} \quad \text{(upper bound).}$$

Example (cont'd): Let's compare the above bound to the *exact* answer.

$$\begin{aligned} & P\left(\left|X-\frac{1}{2}\right| \geq \frac{1}{3}\right) \\ &= 1-P\left(\left|X-\frac{1}{2}\right| < \frac{1}{3}\right) \\ &= 1-P\left(-\frac{1}{3} < X-\frac{1}{2} < \frac{1}{3}\right) \\ &= 1-P\left(\frac{1}{6} < X < \frac{5}{6}\right) \\ &= 1-\int_{1/6}^{5/6} f(x)\,dx \\ &= 1-\frac{2}{3} = 1/3. \end{aligned}$$

(So Chebychev bound was pretty high in comparison.)

Outline

1. Intro / Definitions
2. Discrete Random Variables
3. Continuous Random Variables
4. Cumulative Distribution Functions
5. Great Expectations
6. Functions of a Random Variable
7. Bivariate Random Variables
8. Conditional Distributions
9. Independent Random Variables
10. Conditional Expectation
11. Covariance and Correlation
12. Moment Generating Functions

Functions of a Random Variable

Problem Statement

Discrete Case

Continuous Case

Inverse Transform Theorem

Problem: You have a RV X and you know its pmf/pdf $f(x)$.

Define $Y \equiv h(X)$ (some fn of X).

Find $g(y)$, the pmf/pdf of Y.

Discrete Case: X discrete implies Y discrete implies

$$g(y) = P(Y = y) = P(h(X) = y) = P(\{x | h(x) = y\}) = \sum_{x | h(x) = y} f(x).$$

Example: X is the # of H's in 2 coin tosses. Want pmf for $Y = h(X) = X^3 - X$.

x	0	1	2
$P(X = x)$	$\frac{1}{4}$	$\frac{1}{2}$	$\frac{1}{4}$
$y = x^3 - x$	0	0	6

$g(0) = P(Y = 0) = P(X = 0$ or $1) = 3/4$ and
$g(6) = P(Y = 6) = P(X = 2) = 1/4$.

$$g(y) = \begin{cases} 3/4 & \text{if } y = 0 \\ 1/4 & \text{if } y = 6 \end{cases}$$

Example: X is discrete with

$$f(x) = \begin{cases} 1/8 & \text{if } x = -1 \\ 3/8 & \text{if } x = 0 \\ 1/3 & \text{if } x = 1 \\ 1/6 & \text{if } x = 2 \\ 0 & \text{otherwise} \end{cases}$$

Let $Y = X^2$ (Y can only equal 0,1,4).

$$g(y) = \begin{cases} P(Y = 0) = f(0) = 3/8 \\ P(Y = 1) = f(-1) + f(1) = 11/24 \\ P(Y = 4) = f(2) = 1/6 \\ 0, \ \text{otherwise} \end{cases}$$

Continuous Case: X continuous implies Y can be continuous *or* discrete.

Example: $Y = X^2$ (clearly cts)

Example: $Y = \begin{cases} 0 & \text{if } X < 0 \\ 1 & \text{if } X \geq 0 \end{cases}$ is *not* continuous.

Method: Compute $G(y)$, the cdf of Y.

$$G(y) = P(Y \leq y) = P(h(X) \leq y) = \int_{\{x|h(x)\leq y\}} f(x)\,dx.$$

If $G(y)$ is cts, construct the pdf $g(y)$ by differentiating.

Example: $f(x) = |x|$, $-1 \le x \le 1$.

Find the pdf of the RV $Y = X^2$.

$$G(y) \;=\; P(Y \le y) \;=\; P(X^2 \le y) \;=\; \begin{cases} 0 & \text{if } y \le 0 \\ 1 & \text{if } y \ge 1 \\ (\star) & \text{if } 0 < y < 1 \end{cases},$$

where

$$(\star) \;=\; P(-\sqrt{y} \le X \le \sqrt{y}) \;=\; \int_{-\sqrt{y}}^{\sqrt{y}} |x|\, dx \;=\; y.$$

Thus,

$$G(y) = \begin{cases} 0 & \text{if } y \le 0 \\ 1 & \text{if } y \ge 1 \\ y & \text{if } 0 < y < 1 \end{cases}$$

This implies

$$g(y) = G'(y) = \begin{cases} 0 & \text{if } y \le 0 \text{ or } y \ge 1 \\ 1 & \text{if } 0 < y < 1 \end{cases}$$

This is the U(0,1) distribution!

Example: Suppose $U \sim \mathrm{U}(0,1)$. Find the pdf of $Y = -\ell\mathrm{n}(1-U)$.

$$\begin{aligned} G(y) &= P(Y \le y) \\ &= P(-\ell\mathrm{n}(1-U) \le y) \\ &= P(1-U \ge e^{-y}) \\ &= P(U \le 1-e^{-y}) \\ &= \int_0^{1-e^{-y}} f(u)\,du \\ &= 1-e^{-y} \quad \text{(since } f(u)=1\text{)} \end{aligned}$$

Taking the derivative, we have $g(y) = e^{-y}$, $y > 0$.

Wow! This implies $Y \sim \mathrm{Exp}(\lambda = 1)$.

We can generalize this result…

Inverse Transform Theorem: Suppose X is a cts RV having cdf $F(x)$. Then the *random variable* $F(X) \sim \mathrm{U}(0,1)$.

Proof: Let $Y = F(X)$. Then the cdf of Y is

$$\begin{aligned} G(y) &= P(Y \le y) \\ &= P(F(X) \le y) \\ &= P(X \le F^{-1}(y)) \quad \text{(the cdf is mono. increasing)} \\ &= F(F^{-1}(y)) \quad (F(x) \text{ is the cdf of } X) \\ &= y. \quad \text{Uniform!} \end{aligned}$$

Remark: This is a great theorem, since it applies to all continuous RV's X.

Corollary: $X = F^{-1}(U)$, so you can plug in a U(0,1) RV into the inverse cdf to generate a realization of a RV having X's distribution.

Remark: This is what we did in the example on the previous page. This trick has tremendous applications in simulation.

Outline

1. Intro / Definitions
2. Discrete Random Variables
3. Continuous Random Variables
4. Cumulative Distribution Functions
5. Great Expectations
6. Functions of a Random Variable
7. **Bivariate Random Variables**
8. Conditional Distributions
9. Independent Random Variables
10. Conditional Expectation
11. Covariance and Correlation
12. Moment Generating Functions

Bivariate Random Variables

Discrete Case

Continuous Case

Bivariate cdf's

Marginal Distributions

Now let's look at what happens when you consider 2 RV's simultaneously.

Example: Choose a person at random. Look at height and weight (X, Y). Obviously, X and Y will be related somehow.

Discrete Case

Definition: If X and Y are discrete RV's, then (X, Y) is called a **jointly discrete bivariate RV.**

The joint (or bivariate) pmf is

$$f(x,y) \;=\; P(X = x, Y = y).$$

Properties:

(1) $0 \leq f(x,y) \leq 1$.
(2) $\sum_x \sum_y f(x,y) = 1$.
(3) $A \subseteq \mathbb{R}^2 \Rightarrow P((X,Y) \in A) = \sum\sum_{(x,y)\in A} f(x,y)$.

Example: 3 sox in a box (numbered 1,2,3). Draw 2 sox at random w/o replacement. $X =$ # of first sock, $Y =$ # of second sock. The joint pmf $f(x, y)$ is

	$X = 1$	$X = 2$	$X = 3$	$P(Y = y)$
$Y = 1$	0	1/6	1/6	1/3
$Y = 2$	1/6	0	1/6	1/3
$Y = 3$	1/6	1/6	0	1/3
$P(X = x)$	1/3	1/3	1/3	1

$f_X(x) \equiv P(X = x)$ is the “marginal” distribution of X.
$f_Y(y) \equiv P(Y = y)$ is the “marginal” distribution of Y.

By the law of total probability,

$$P(X=1) \;=\; \sum_{y=1}^{3} P(X=1, Y=y) \;=\; 1/3.$$

In addition,

$$\begin{aligned} & P(X \geq 2, Y \geq 2) \\ & \quad = \sum_{x \geq 2} \sum_{y \geq 2} f(x,y) \\ & \quad = f(2,2) + f(2,3) + f(3,2) + f(3,3) \\ & \quad = 0 + 1/6 + 1/6 + 0 \;=\; 1/3. \end{aligned}$$

Continuous Case

Definition: If X and Y are cts RV's, then (X, Y) is a **jointly cts RV** if there exists a function $f(x, y)$ such that

(1) $f(x, y) \geq 0$, $\forall x, y$.
(2) $\int\int_{\mathbb{R}^2} f(x, y)\, dx\, dy = 1$.
(3) $P(A) = P((X, Y) \in A) = \int\int_A f(x, y)\, dx\, dy$.

In this case, $f(x, y)$ is called the **joint pdf**.

If $A \subseteq \mathbb{R}^2$, then $P(A)$ is the volume between $f(x, y)$ and A.

Think of

$$f(x, y)\, dx\, dy \;\approx\; P(x < X < x + dx, y < Y < y + dy).$$

Easy to see how all of this generalizes the 1-dimensional pdf, $f(x)$.

Easy Example: Choose a point at random in the interior of the circle inscribed in the unit square, e.g., $C \equiv (x - \frac{1}{2})^2 + (y - \frac{1}{2})^2 = \frac{1}{4}$. Find the pdf of (X, Y).

Since the area of the circle is $\pi/4$,

$$f(x,y) = \begin{cases} 4/\pi & \text{if } (x,y) \in C \\ 0 & \text{otherwise} \end{cases}$$

Application: Toss n darts randomly into the unit square. The probability that any individual dart will land in the circle is $\pi/4$. It stands to reason that the proportion of darts, $\hat{p}_n$, that land in the circle will be approximately $\pi/4$. So you can use $4\hat{p}_n$ to estimate π!

Example: Suppose that

$$f(x,y) = \begin{cases} 4xy & \text{if } 0 \le x \le 1,\ 0 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Find the prob (volume) of the region $0 \le y \le 1 - x^2$.

$$\begin{aligned} V &= \int_0^1 \int_0^{1-x^2} 4xy \, dy \, dx \\ &= \int_0^1 \int_0^{\sqrt{1-y}} 4xy \, dx \, dy \\ &= 1/3. \end{aligned}$$

Moral: Be careful with limits!

Bivariate cdf's

Definition: The **joint (bivariate) cdf** of X and Y is $F(x,y) \equiv P(X \le x, Y \le y)$, for all x, y.

$$F(x,y) = \begin{cases} \sum\sum_{s \le x, t \le y} f(s,t) & \text{discrete} \\ \\ \int_{-\infty}^{y}\int_{-\infty}^{x} f(s,t)\,ds\,dt & \text{continuous} \end{cases}$$

Properties:

$F(x,y)$ is non-decreasing in both x and y.

$\lim_{x\to-\infty} F(x,y) = \lim_{y\to-\infty} F(x,y) = 0$.

$\lim_{x\to\infty} F(x,y) = F_Y(y) = P(Y \le y)$
$\lim_{y\to\infty} F(x,y) = F_X(x) = P(X \le x)$.

$\lim_{x\to\infty} \lim_{y\to\infty} F(x,y) = 1$.

$F(x,y)$ is cts from the right in both x and y.

Going from cdf's to pdf's (continuous case).

1-dimension: $f(x) = F'(x) = \frac{d}{dx} \int_{-\infty}^{x} f(t)\,dt$.

2-D: $f(x,y) = \frac{\partial^2}{\partial x \partial y} F(x,y) = \frac{\partial^2}{\partial x \partial y} \int_{-\infty}^{x} \int_{-\infty}^{y} f(s,t)\,dt\,ds$.

Example: Suppose

$$F(x,y) = \begin{cases} 1 - e^{-x} - e^{-y} + e^{-(x+y)} & \text{if } x \ge 0,\ y \ge 0 \\ 0 & \text{if } x < 0 \text{ or } y < 0 \end{cases}$$

The "marginal" cdf of X is

$$F_X(x) = \lim_{y\to\infty} F(x,y) = \begin{cases} 1 - e^{-x} & \text{if } x \ge 0 \\ 0 & \text{if } x < 0 \end{cases}$$

The joint pdf is

$$\begin{aligned} f(x,y) &= \frac{\partial^2}{\partial x \partial y} F(x,y) \\ &= \frac{\partial}{\partial y}(e^{-x} - e^{-y}e^{-x}) \\ &= \begin{cases} e^{-(x+y)} & \text{if } x \ge 0,\ y \ge 0 \\ 0 & \text{if } x < 0 \text{ or } y < 0 \end{cases} \end{aligned}$$

Marginal Distributions — Distrns of X and Y.

Example (discrete case): $f(x,y) = P(X = x, Y = y)$.

	$X=1$	$X=2$	$X=3$	$P(Y=y)$
$Y=4$	0.01	0.07	0.12	0.2
$Y=6$	0.29	0.03	0.48	0.8
$P(X=x)$	0.3	0.1	0.6	1

By total probability,

$$P(X=1) = P(X=1, Y = \text{ any \#}) = 0.3.$$

Definition: If X and Y are jointly discrete, then the **marginal pmf's** of X and Y are, respectively,

$$f_X(x) \;=\; P(X=x) \;=\; \sum_y f(x,y)$$

and

$$f_Y(y) \;=\; P(Y=y) \;=\; \sum_x f(x,y)$$

Example (discrete case): $f(x,y) = P(X=x, Y=y)$.

	$X=1$	$X=2$	$X=3$	$P(Y=y)$
$Y=4$	0.06	0.02	0.12	0.2
$Y=6$	0.24	0.08	0.48	0.8
$P(X=x)$	0.3	0.1	0.6	1

Hmmm.... Compared to the last example, this has the *same marginals* but *different joint* distribution!

Definition: If X and Y are jointly cts, then the **marginal pdf's** of X and Y are, respectively,

$$f_X(x) \;=\; \int_{\mathbb{R}} f(x,y)\,dy \quad \text{and} \quad f_Y(y) \;=\; \int_{\mathbb{R}} f(x,y)\,dx.$$

Example:

$$f(x,y) \;=\; \begin{cases} e^{-(x+y)} & \text{if } x \geq 0,\, y \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Then the marginal pdf of X is

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy = \int_0^\infty e^{-(x+y)}\,dy = \begin{cases} e^{-x} & \text{if } x \geq 0 \\ 0 & \text{otherwise} \end{cases}$$

Example:

$$f(x,y) = \begin{cases} \frac{21}{4}x^2y & \text{if } x^2 \le y \le 1 \\ 0 & \text{otherwise} \end{cases}$$

Note funny limits where the pdf is positive, i.e., $x^2 \le y \le 1$.

$$f_X(x) = \int_{\mathbb{R}} f(x,y)\,dy = \int_{x^2}^{1} \frac{21}{4}x^2y\,dy = \frac{21}{8}x^2(1-x^4), \quad \text{if } -1 \le x \le 1$$

$$f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx = \int_{-\sqrt{y}}^{\sqrt{y}} \frac{21}{4}x^2y\,dx = \frac{7}{2}y^{5/2}, \quad \text{if } 0 \le y \le 1$$

Outline

1. Intro / Definitions
2. Discrete Random Variables
3. Continuous Random Variables
4. Cumulative Distribution Functions
5. Great Expectations
6. Functions of a Random Variable
7. Bivariate Random Variables
8. **Conditional Distributions**
9. Independent Random Variables
10. Conditional Expectation
11. Covariance and Correlation
12. Moment Generating Functions

Conditional Distributions

Recall conditional probability: $P(A|B) = P(A \cap B)/P(B)$ if $P(B) > 0$.

Suppose that X and Y are jointly discrete RV's. Then if $P(Y = y) > 0$,

$$P(X = x|Y = y) \;=\; \frac{P(X = x \cap Y = y)}{P(Y = y)} \;=\; \frac{f(x,y)}{f_Y(y)}$$

$P(X = x|Y = 2)$ defines the probabilities on X given that $Y = 2$.

Definition: If $f_Y(y) > 0$, then $f_{X|Y}(x|y) \equiv \frac{f(x,y)}{f_Y(y)}$ is the **conditional pmf/pdf of** X **given** $Y = y$.

Remark: Usually just write $f(x|y)$ instead of $f_{X|Y}(x|y)$.

Remark: Of course, $f_{Y|X}(y|x) = f(y|x) = \frac{f(x,y)}{f_X(x)}$.

Discrete Example: $f(x,y) = P(X=x, Y=y)$.

	$X=1$	$X=2$	$X=3$	$f_Y(y)$
$Y=4$	0.01	0.07	0.12	0.2
$Y=6$	**0.29**	**0.03**	**0.48**	**0.8**
$f_X(x)$	0.3	0.1	0.6	1

Then, for example,

$$f(x|y=6) \;=\; \frac{f(x,6)}{f_Y(6)} \;=\; \frac{f(x,6)}{0.8} \;=\; \begin{cases} \frac{29}{80} & \text{if } x=1 \\ \frac{3}{80} & \text{if } x=2 \\ \frac{48}{80} & \text{if } x=3 \\ 0 & \text{otherwise} \end{cases}$$

Old Cts Example:

$$f(x,y) = \frac{21}{4}x^2y, \quad \text{if } x^2 \le y \le 1$$

$$f_X(x) = \frac{21}{8}x^2(1-x^4), \quad \text{if } -1 \le x \le 1$$

$$f_Y(y) = \frac{7}{2}y^{5/2}, \quad \text{if } 0 \le y \le 1$$

Then, for example,

$$f(y|\frac{1}{2}) = \frac{f(\frac{1}{2},y)}{f_X(\frac{1}{2})} = \frac{\frac{21}{4}\cdot\frac{1}{4}y}{\frac{21}{8}\cdot\frac{1}{4}\cdot(1-\frac{1}{16})} = \frac{32}{15}y, \quad \text{if } \tfrac{1}{4} \le y \le 1.$$

More generally,

$$\begin{aligned} f(y|x) &= \frac{f(x,y)}{f_X(x)} \\ &= \frac{\frac{21}{4}x^2 y}{\frac{21}{8}x^2(1-x^4)}, \quad \text{if } x^2 \le y \le 1 \\ &= \frac{2y}{1-x^4}, \quad \text{if } x^2 \le y \le 1. \end{aligned}$$

Note: $2/(1-x^4)$ is a constant with respect to y, and we can check to see that $f(y|x)$ is a legit condl pdf:

$$\int_{x^2}^{1} \frac{2y}{1-x^4}\, dy \;=\; 1.$$

Typical Problem: Given $f_X(x)$ and $f(y|x)$, find $f_Y(y)$.

Steps: (1) $f(x,y) = f_X(x)f(y|x)$

(2) $f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx$.

Example: $f_X(x) = 2x, \;\; 0 < x < 1$.

Given $X = x$, suppose that $Y|x \sim \mathrm{U}(0,x)$. Now find $f_Y(y)$.

Solution: $Y|x \sim \text{U}(0,x) \;\Rightarrow\; f(y|x) = 1/x,\;\; 0 < y < x$. So

$$\begin{aligned} f(x,y) &= f_X(x) f(y|x) \\ &= 2x \cdot \frac{1}{x}, \;\; \text{if } 0 < x < 1 \text{ and } 0 < y < x \\ &= 2, \;\; \text{if } 0 < y < x < 1. \end{aligned}$$

Thus,

$$f_Y(y) = \int_{\mathbb{R}} f(x,y)\,dx = \int_y^1 2\,dx = 2(1-y), \;\; 0 < y < 1.$$

Outline

1. Intro / Definitions
2. Discrete Random Variables
3. Continuous Random Variables
4. Cumulative Distribution Functions
5. Great Expectations
6. Functions of a Random Variable
7. Bivariate Random Variables
8. Conditional Distributions
9. **Independent Random Variables**
10. Conditional Expectation
11. Covariance and Correlation
12. Moment Generating Functions

Independence

Recall that two events are independent if $P(A \cap B) = P(A)P(B)$.

Then

$$P(A|B) \;=\; \frac{P(A \cap B)}{P(B)} \;=\; \frac{P(A)P(B)}{P(B)} \;=\; P(A).$$

And similarly, $P(B|A) = P(B)$.

Now want to define independence for RV's, i.e., the outcome of X doesn't influence the outcome of Y.

Definition: X and Y are **independent** RV's if, for all x and y,

$$f(x, y) \;=\; f_X(x) f_Y(y).$$

Equivalent definitions:

$$F(x,y) \;=\; F_X(x)F_Y(y),\;\; \forall x,y$$

or

$$P(X \le x, Y \le y) \;=\; P(X \le x)P(Y \le y),\;\; \forall x,y$$

If X and Y aren't indep, then they're **dependent.**

Theorem: X and Y are indep if and only if $f(y|x) = f_Y(y)\; \forall x,y$.

Proof:

$$f(y|x) \;=\; \frac{f(x,y)}{f_X(x)} \;=\; \frac{f_X(x)f_Y(y)}{f_X(x)} \;=\; f_Y(y).$$

Similarly, X and Y indep implies $f(x|y) = f_X(x)$.

Example (discrete): $f(x,y) = P(X = x, Y = y)$.

	$X = 1$	$X = 2$	$f_Y(y)$
$Y = 2$	0.12	0.28	0.4
$Y = 3$	0.18	0.42	0.6
$f_X(x)$	0.3	0.7	1

X and Y are indep since $f(x,y) = f_X(x)f_Y(y)$, $\forall x, y$.

Example (cts): Suppose $f(x,y) = 6xy^2$, $0 \le x \le 1$, $0 \le y \le 1$.
After some work (which can be avoided by the next theorem), we can derive

$$\begin{aligned} f_X(x) &= 2x, \text{ if } 0 \le x \le 1, \text{ and} \\ f_Y(y) &= 3y^2, \text{ if } 0 \le y \le 1. \end{aligned}$$

X and Y are indep since $f(x,y) = f_X(x) f_Y(y)$, $\forall x, y$.

Easy way to tell if X and Y are indep...

Theorem: X and Y are indep iff $f(x,y) = a(x)b(y)$, $\forall x, y$, for some functions $a(x)$ and $b(y)$ (not necessarily pdf's).

So if $f(x,y)$ factors into separate functions of x and y, then X and Y are indep.

Example: $f(x,y) = 6xy^2$, $0 \le x \le 1$, $0 \le y \le 1$. Take

$$a(x) \;=\; 6x, \;\; 0 \le x \le 1, \text{ and } b(y) \;=\; y^2, \;\; 0 \le y \le 1.$$

Thus, X and Y are indep (as above).

Example: $f(x,y) = \frac{21}{4}x^2y$, $x^2 \le y \le 1$.

"Funny" (non-rectangular) limits make factoring into marginals impossible. Thus, X and Y are *not* indep.

Example: $f(x,y) = \frac{c}{x+y}$, $1 \le x \le 2$, $1 \le y \le 3$.

Can't factor $f(x,y)$ into fn's of x and y separately. Thus, X and Y are *not* indep.

Now that we can figure out if X and Y are indep, what can we do with that knowledge?

Consequences of Independence

Definition/Theorem (another Unconscious Statistician): Let $h(X,Y)$ be a fn of the RV's X and Y. Then

$$\mathrm{E}[h(X,Y)] \;=\; \begin{cases} \sum_x \sum_y h(x,y) f(x,y) & \text{discrete} \\ \int_{\mathbb{R}} \int_{\mathbb{R}} h(x,y) f(x,y)\, dx\, dy & \text{continuous} \end{cases}$$

Theorem: *Whether or not* X and Y are indep,

$$\mathrm{E}[X+Y] \;=\; \mathrm{E}[X] + \mathrm{E}[Y].$$

Proof (cts case):

$$
\begin{aligned}
\mathrm{E}[X+Y] &= \int_{\mathbb{R}}\int_{\mathbb{R}} (x+y) f(x,y)\,dx\,dy \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} x f(x,y)\,dx\,dy + \int_{\mathbb{R}}\int_{\mathbb{R}} y f(x,y)\,dx\,dy \\
&= \int_{\mathbb{R}} x \int_{\mathbb{R}} f(x,y)\,dy\,dx + \int_{\mathbb{R}} y \int_{\mathbb{R}} f(x,y)\,dx\,dy \\
&= \int_{\mathbb{R}} x f_X(x)\,dx + \int_{\mathbb{R}} y f_Y(y)\,dy \\
&= \mathrm{E}[X] + \mathrm{E}[Y].
\end{aligned}
$$

Can generalize this result to more than two RV's.

Corollary: If $X_1, X_2, \ldots, X_n$ are RV's, then

$$\mathrm{E}\Big[\sum_{i=1}^{n} X_i\Big] \;=\; \sum_{i=1}^{n} \mathrm{E}[X_i].$$

Proof: Induction.

Theorem: If X and Y are *indep*, then $\mathrm{E}[XY] = \mathrm{E}[X]\mathrm{E}[Y]$.

Proof (cts case):

$$\begin{aligned}
\mathrm{E}[XY] &= \int_{\mathbb{R}}\int_{\mathbb{R}} xyf(x,y)\,dx\,dy \\
&= \int_{\mathbb{R}}\int_{\mathbb{R}} xyf_X(x)f_Y(y)\,dx\,dy \quad \text{(}X\text{ and }Y\text{ are indep)} \\
&= \left(\int_{\mathbb{R}} xf_X(x)\,dx\right)\left(\int_{\mathbb{R}} yf_Y(y)\,dy\right) \\
&= \mathrm{E}[X]\mathrm{E}[Y].
\end{aligned}$$

Remark: The above theorem is *not* necessarily true if X and Y are *dependent*. See the upcoming discussion on covariance.

Theorem: If X and Y are *indep*, then

$$\text{Var}(X+Y) \;=\; \text{Var}(X)+\text{Var}(Y).$$

Remark: The assumption of independence really is important here. If X and Y aren't independent, then the result might not hold.

Corollary: If $X_1, X_2, \ldots, X_n$ are *indep* RV's, then

$$\text{Var}\left(\sum_{i=1}^{n} X_i\right) \;=\; \sum_{i=1}^{n} \text{Var}(X_i).$$

Proof: Induction.

Proof:

$$
\begin{aligned}
&\mathrm{Var}(X+Y) \\
&= \mathrm{E}[(X+Y)^2] - (\mathrm{E}[X+Y])^2 \\
&= \mathrm{E}[X^2 + 2XY + Y^2] - (\mathrm{E}[X] + \mathrm{E}[Y])^2 \\
&= \mathrm{E}[X^2] + 2\mathrm{E}[XY] + \mathrm{E}[Y^2] \\
&\qquad -(\mathrm{E}[X])^2 - 2\mathrm{E}[X]\mathrm{E}[Y] - (\mathrm{E}[Y])^2 \\
&= \mathrm{E}[X^2] + 2\mathrm{E}[X]\mathrm{E}[Y] + \mathrm{E}[Y^2] \\
&\qquad -(\mathrm{E}[X])^2 - 2\mathrm{E}[X]\mathrm{E}[Y] - (\mathrm{E}[Y])^2 \\
&= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 + \mathrm{E}[Y^2] - (\mathrm{E}[Y])^2 \\
&= \mathrm{Var}(X) + \mathrm{Var}(Y).
\end{aligned}
$$

Random Samples

Definition: $X_1, X_2, \ldots, X_n$ form a **random sample** if

- X_i's are all *independent*.
- Each X_i has the same pmf/pdf $f(x)$.

Notation: $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} f(x)$ ("indep and identically distributed")

Example/Theorem: Suppose $X_1, \ldots, X_n \stackrel{\text{iid}}{\sim} f(x)$ with $\mathrm{E}[X_i] = \mu$ and $\mathrm{Var}(X_i) = \sigma^2$. Define the **sample mean** as

$$\bar{X} \equiv \frac{1}{n} \sum_{i=1}^{n} X_i.$$

Then

$$\mathrm{E}[\bar{X}] = \mathrm{E}\left[\frac{1}{n} \sum_{i=1}^{n} X_i\right] = \frac{1}{n} \sum_{i=1}^{n} \mathrm{E}[X_i] = \frac{1}{n} \sum_{i=1}^{n} \mu = \mu.$$

So the mean of $\bar{X}$ is the same as the mean of X_i.

Meanwhile,...

$$\begin{aligned}
\text{Var}(\bar{X}) &= \text{Var}\left(\frac{1}{n}\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\text{Var}\left(\sum_{i=1}^{n} X_i\right) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\text{Var}(X_i) \quad (X_i\text{'s indep}) \\
&= \frac{1}{n^2}\sum_{i=1}^{n}\sigma^2 \;=\; \sigma^2/n.
\end{aligned}$$

So the mean of $\bar{X}$ is the same as the mean of X_i, but the *variance decreases*! This makes $\bar{X}$ a great *estimator* for μ (which is usually unknown in practice), and the result is referred to as the **Law of Large Numbers**. Stay tuned.

Outline

1. Intro / Definitions
2. Discrete Random Variables
3. Continuous Random Variables
4. Cumulative Distribution Functions
5. Great Expectations
6. Functions of a Random Variable
7. Bivariate Random Variables
8. Conditional Distributions
9. Independent Random Variables
10. Conditional Expectation
11. Covariance and Correlation
12. Moment Generating Functions

Conditional Expectation

Usual definition of expectation. E.g., what's the avg weight of a male?

$$\mathrm{E}[Y] = \begin{cases} \sum_y y f(y) & \text{discrete} \\ \int_{\mathbb{R}} y f(y)\, dy & \text{continuous} \end{cases}$$

Now suppose we're interested in the avg weight of a 6'6" tall male.

$f(y|x)$ is the conditional pdf/pmf of Y given $X = x$.

Definition: The **conditional expectation** of Y given $X = x$ is

$$\mathrm{E}[Y|X = x] \equiv \begin{cases} \sum_y y f(y|x) & \text{discrete} \\ \int_{\mathbb{R}} y f(y|x)\, dy & \text{continuous} \end{cases}$$

Note that $\mathrm{E}[Y|X = x]$ is a function of x.

Discrete Example:

$f(x,y)$	$X=0$	$X=3$	$f_Y(y)$
$Y=2$	0.11	**0.34**	0.45
$Y=5$	0.00	**0.05**	0.05
$Y=10$	0.29	**0.21**	0.50
$f_X(x)$	0.40	**0.60**	1

The *unconditional* expectation is $\mathrm{E}[Y]=\sum_y y f_Y(y)=6.15$. But conditional on $X=3$, we have

$$f(y|x=3) \;=\; \frac{f(3,y)}{f_X(3)} \;=\; \frac{f(3,y)}{0.60} \;=\; \begin{cases} \frac{34}{60} & \text{if } y=2 \\ \frac{5}{60} & \text{if } y=5 \\ \frac{21}{60} & \text{if } y=10 \end{cases}$$

So the expectation conditional on $X=3$ is

$$\mathrm{E}[Y|X=3] \;=\; \sum_y y f(y|3) \;=\; 2\Big(\frac{34}{60}\Big)+5\Big(\frac{5}{60}\Big)+10\Big(\frac{21}{60}\Big) \;=\; 5.05.$$

Old Cts Example:

$$f(x,y) \;=\; \frac{21}{4}x^2y, \quad \text{if } x^2 \le y \le 1.$$

Recall that

$$f(y|x) \;=\; \frac{2y}{1-x^4} \quad \text{if } x^2 \le y \le 1.$$

Thus,

$$\mathrm{E}[Y|x] \;=\; \int_{\mathbb{R}} y f(y|x)\,dy \;=\; \frac{2}{1-x^4}\int_{x^2}^{1} y^2\,dy \;=\; \frac{2}{3}\cdot\frac{1-x^6}{1-x^4}.$$

Thus, e.g., $\mathrm{E}[Y|X=0.5] = \frac{2}{3}\cdot\frac{63}{64}/\frac{15}{16} = 0.70$.

Theorem (double expectations): $\mathrm{E}[\mathrm{E}(Y|X)] = \mathrm{E}[Y]$.

Remarks: Yikes, what the heck is this!?

The expected value (averaged over all X's) of the conditional expected value (of $Y|X$) is the plain old expected value (of Y).

Think of the outside exp value as the exp value of $h(X) = \mathrm{E}(Y|X)$. Then the Law of the Unconscious Statistician miraculously gives us $\mathrm{E}[Y]$.

Believe it or not, sometimes it's easier to calculate $\mathrm{E}[Y]$ indirectly by using our double expectation trick.

Proof (cts case): By the Unconscious Statistician,

$$\begin{aligned}
\mathrm{E}[\mathrm{E}(Y|X)] &= \int_{\mathbb{R}} \mathrm{E}(Y|x) f_X(x)\,dx \\
&= \int_{\mathbb{R}} \left(\int_{\mathbb{R}} y f(y|x)\,dy \right) f_X(x)\,dx \\
&= \int_{\mathbb{R}} \int_{\mathbb{R}} y f(y|x) f_X(x)\,dx\,dy \\
&= \int_{\mathbb{R}} y \int_{\mathbb{R}} f(x,y)\,dx\,dy \\
&= \int_{\mathbb{R}} y f_Y(y)\,dy \;=\; \mathrm{E}[Y].
\end{aligned}$$

Old Example: Suppose $f(x,y) = \frac{21}{4}x^2y$, if $x^2 \le y \le 1$. Find $\mathrm{E}[Y]$ **two ways.**

By previous examples, we know that

$$\begin{aligned} f_X(x) &= \frac{21}{8}x^2(1-x^4), \quad \text{if } -1 \le x \le 1 \\ f_Y(y) &= \frac{7}{2}y^{5/2}, \quad \text{if } 0 \le y \le 1 \\ \mathrm{E}[Y|x] &= \frac{2}{3} \cdot \frac{1-x^6}{1-x^4}. \end{aligned}$$

Solution #1 (old, boring way):

$$\mathrm{E}[Y] \;=\; \int_{\mathbb{R}} y f_Y(y)\,dy \;=\; \int_0^1 \frac{7}{2} y^{7/2}\,dy \;=\; \frac{7}{9}.$$

Solution #2 (new, exciting way):

$$\begin{aligned}
\mathrm{E}[Y] &= \mathrm{E}[\mathrm{E}(Y|X)] \\
&= \int_{\mathbb{R}} \mathrm{E}(Y|x) f_X(x)\,dx \\
&= \int_{-1}^{1} \left(\frac{2}{3}\cdot\frac{1-x^6}{1-x^4}\right)\left(\frac{21}{8}x^2(1-x^4)\right)dx \;=\; \frac{7}{9}.
\end{aligned}$$

Notice that both answers are the same (good)!

Example: An alternative way to calculate the mean of the Geom(p).

Let $Y \sim \mathrm{Geom}(p)$, e.g., Y could be the number of coin flips before H appears, where $P(H) = p$. Further, let

$$X = \begin{cases} 1 & \text{if first flip is H} \\ 0 & \text{otherwise} \end{cases}.$$

Based on the result X of the first step, we have

$$\begin{aligned} \mathrm{E}[Y] &= \mathrm{E}[\mathrm{E}(Y|X)] \\ &= \sum_x \mathrm{E}(Y|x) f_X(x) \\ &= \mathrm{E}(Y|X=0)P(X=0) + \mathrm{E}(Y|X=1)P(X=1) \\ &= (1+\mathrm{E}[Y])(1-p) + 1(p). \quad \text{(why?)} \end{aligned}$$

Solving, we get $\mathrm{E}[Y] = 1/p$ (which is indeed the correct answer!)

Theorem (expectation of the sum of a random number of RV's):

Suppose that $X_1, X_2, \ldots$ are independent RV's, all with the same mean. Also suppose that N is a nonnegative, integer-valued RV, that's independent of the X_i's. Then

$$\mathrm{E}\left(\sum_{i=1}^{N} X_i\right) = \mathrm{E}[N]\mathrm{E}[X_1].$$

Remark: You have to be very careful here. In particular, note that $\mathrm{E}(\sum_{i=1}^{N} X_i) \neq N\mathrm{E}[X_1]$, since the LHS is a number and the RHS is random.

Proof: By double expectation and the fact that N is indep of the X_i's,

$$
\begin{aligned}
\mathrm{E}\left(\sum_{i=1}^{N} X_i\right) &= \mathrm{E}\left[\mathrm{E}\left(\sum_{i=1}^{N} X_i \Big| N\right)\right] \\
&= \sum_{n=1}^{\infty} \mathrm{E}\left(\sum_{i=1}^{N} X_i \Big| N=n\right) P(N=n) \\
&= \sum_{n=1}^{\infty} \mathrm{E}\left(\sum_{i=1}^{n} X_i \Big| N=n\right) P(N=n) \\
&= \sum_{n=1}^{\infty} \mathrm{E}\left(\sum_{i=1}^{n} X_i\right) P(N=n) \\
&= \sum_{n=1}^{\infty} n\mathrm{E}[X_1] P(N=n) \\
&= \mathrm{E}[X_1] \sum_{n=1}^{\infty} nP(N=n). \quad \square
\end{aligned}
$$

Example: Suppose the number of times we roll a die is $N \sim \text{Pois}(10)$. If X_i denotes the value of the ith toss, then the expected total of all of the rolls is

$$\mathrm{E}\left(\sum_{i=1}^{N} X_i\right) = \mathrm{E}[N]\mathrm{E}[X_1] = 10(3.5) = 35. \quad \Box$$

Theorem: Under the same conditions as before,

$$\mathrm{Var}\left(\sum_{i=1}^{N} X_i\right) = \mathrm{E}[N]\mathrm{Var}(X_1) + (\mathrm{E}[X_1])^2\mathrm{Var}(N).$$

Proof: See, for instance, Ross. $\Box$

Computing Probabilities by Conditioning

Let A be some event, and define the RV Y as:

$$Y = \begin{cases} 1 & \text{if } A \text{ occurs} \\ 0 & \text{otherwise} \end{cases}.$$

Then

$$\mathrm{E}[Y] = \sum_y y f_Y(y) = P(Y=1) = P(A).$$

Similarly, for any RV X, we have

$$\begin{aligned} \mathrm{E}[Y|X=x] &= \sum_y y f_Y(y|x) \\ &= P(Y=1|X=x) \\ &= P(A|X=x). \end{aligned}$$

Further, since $\mathrm{E}[Y] = \mathrm{E}[\mathrm{E}(Y|X)]$, we have

$$\begin{aligned} P(A) &= \mathrm{E}[Y] \\ &= \mathrm{E}[\mathrm{E}(Y|X)] \\ &= \int_{\mathbb{R}} \mathrm{E}[Y|x] dF_X(x) \\ &= \int_{\mathbb{R}} P(A|X = x) dF_X(x). \end{aligned}$$

Example/Theorem: If X and Y are independent continuous RV's, then

$$P(Y < X) \;=\; \int_{\mathbb{R}} F_Y(x) f_X(x)\, dx,$$

where $F_Y(\cdot)$ is the c.d.f. of Y and $f_X(\cdot)$ is the p.d.f. of X.

Proof: (Actually, there are many proofs.) Let the event $A = \{Y < X\}$. Then

$$\begin{aligned} P(Y < X) &= \int_{\mathbb{R}} P(Y < X | X = x) f_X(x)\, dx \\ &= \int_{\mathbb{R}} P(Y < x | X = x) f_X(x)\, dx \\ &= \int_{\mathbb{R}} P(Y < x) f_X(x)\, dx \end{aligned}$$

(since X, Y are indep). □

Example: If $X \sim \text{Exp}(\mu)$ and $Y \sim \text{Exp}(\lambda)$ are independent RV's. Then

$$
\begin{aligned}
P(Y < X) &= \int_{\mathbb{R}} F_Y(x) f_X(x)\, dx \\
&= \int_0^\infty (1 - e^{-\lambda x}) \mu e^{-\mu x}\, dx \\
&= \frac{\lambda}{\lambda + \mu}. \quad \Box
\end{aligned}
$$

Example/Theorem: If X and Y are independent continuous RV's, then

$$P(X+Y<a) \;=\; \int_{\mathbb{R}} F_Y(a-x) f_X(x)\,dx,$$

where $F_Y(\cdot)$ is the c.d.f. of Y and $f_X(\cdot)$ is the p.d.f. of X. The quantity $X+Y$ is called a *convolution*.

Proof:

$$\begin{aligned} P(X+Y<a) &= \int_{\mathbb{R}} P(X+Y<a|X=x) f_X(x)\,dx \\ &= \int_{\mathbb{R}} P(Y<a-x|X=x) f_X(x)\,dx \\ &= \int_{\mathbb{R}} P(Y<a-x) f_X(x)\,dx \\ & \quad \text{(since } X, Y \text{ are indep).} \quad \Box \end{aligned}$$

Example: Suppose $X, Y \overset{\text{iid}}{\sim} \text{Exp}(\lambda)$. Note that

$$F_Y(a-x) = \begin{cases} 1 - e^{-\lambda(a-x)} & \text{if } a - x \geq 0 \text{ and } x \geq 0 \\ & \quad \text{(i.e., } 0 \leq x \leq a\text{)} \\ 0 & \text{if otherwise} \end{cases}$$

$$\begin{aligned} P(X+Y<a) &= \int_{\mathbb{R}} F_Y(a-x) f_X(x)\, dx \\ &= \int_0^a (1 - e^{-\lambda(a-x)})\lambda e^{-\lambda x}\, dx \\ &= 1 - e^{-\lambda a} - \lambda a e^{-\lambda a}, \quad \text{if } a \geq 0. \end{aligned}$$

$$\frac{d}{da} P(X+Y<a) = \lambda^2 a e^{-\lambda a}, \quad a \geq 0.$$

This implies that $X + Y \sim \text{Gamma}(2, \lambda)$. $\Box$

And Now, A Word From Our Sponsor. . .

Congratulations! You are now done with the most difficult module of the course!

Things will get easier from here (I hope)!

Outline

1. Intro / Definitions
2. Discrete Random Variables
3. Continuous Random Variables
4. Cumulative Distribution Functions
5. Great Expectations
6. Functions of a Random Variable
7. Bivariate Random Variables
8. Conditional Distributions
9. Independent Random Variables
10. Conditional Expectation
11. Covariance and Correlation
12. Moment Generating Functions

Covariance and Correlation

These are measures used to define the degree of association between X and Y if they don't happen to be indep.

Definition: The **covariance** between X and Y is

$$\mathrm{Cov}(X,Y) \;\equiv\; \sigma_{XY} \;\equiv\; \mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])].$$

Remark: $\mathrm{Cov}(X,X) = \mathrm{E}[(X-\mathrm{E}[X])^2] = \mathrm{Var}(X)$.

If X and Y have positive covariance, then X and Y move "in the same direction." Think height and weight.

If X and Y have negative covariance, then X and Y move "in opposite directions." Think snowfall and temperature.

Theorem (easier way to calculate Cov):

$$\mathrm{Cov}(X,Y) \;=\; \mathrm{E}[XY]-\mathrm{E}[X]\mathrm{E}[Y].$$

Proof:

$$\begin{aligned}
\mathrm{Cov}(X,Y) &= \mathrm{E}[(X-\mathrm{E}[X])(Y-\mathrm{E}[Y])] \\
&= \mathrm{E}\Big[XY - X\mathrm{E}[Y] - Y\mathrm{E}[X] + \mathrm{E}[X]\mathrm{E}[Y]\Big] \\
&= \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] - \mathrm{E}[Y]\mathrm{E}[X] + \mathrm{E}[X]\mathrm{E}[Y] \\
&= \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y].
\end{aligned}$$

Theorem: X and Y indep implies $\text{Cov}(X,Y) = 0$.

Proof:

$$\begin{aligned}
\text{Cov}(X,Y) &= \text{E}[XY] - \text{E}[X]\text{E}[Y] \\
&= \text{E}[X]\text{E}[Y] - \text{E}[X]\text{E}[Y] \quad \text{(}X\text{, }Y\text{ indep)} \\
&= 0.
\end{aligned}$$

Danger Will Robinson: $\text{Cov}(X, Y) = 0$ *does not imply* X and Y are indep!!

Example: Suppose $X \sim \text{U}(-1, 1)$ and $Y = X^2$ (so X and Y are clearly *dependent*).

But

$$\text{E}[X] \;=\; \int_{-1}^{1} x \cdot \frac{1}{2}\, dx \;=\; 0 \;\text{ and}$$

$$\text{E}[XY] \;=\; \text{E}[X^3] \;=\; \int_{-1}^{1} x^3 \cdot \frac{1}{2}\, dx \;=\; 0,$$

so

$$\text{Cov}(X, Y) \;=\; \text{E}[XY] - \text{E}[X]\text{E}[Y] \;=\; 0.$$

Definition: The **correlation** between X and Y is

$$\rho \;=\; \text{Corr}(X,Y) \;\equiv\; \frac{\text{Cov}(X,Y)}{\sqrt{\text{Var}(X)\text{Var}(Y)}} \;=\; \frac{\sigma_{XY}}{\sigma_X \sigma_Y}.$$

Remark: Cov has “square” units; corr is unitless.

Corollary: X, Y indep implies $\rho = 0$.

Theorem: It can be shown that $-1 \le \rho \le 1$.

$\rho \approx 1$ is “high” corr
$\rho \approx 0$ is “low” corr
$\rho \approx -1$ is “high” negative corr.

Example: Height is *highly* correlated with weight.
Temperature on Mars has *low* corr with IBM stock price.

Anti-UGA Example: Suppose X is the avg GPA of a UGA student, and Y is his IQ. Here's the joint pmf.

$f(x,y)$	$X=2$	$X=3$	$X=4$	$f_Y(y)$
$Y=40$	0.0	0.2	0.1	0.3
$Y=50$	0.15	0.1	0.05	0.3
$Y=60$	0.3	0.0	0.1	0.4
$f_X(x)$	0.45	0.3	0.25	1

$$\begin{aligned} \mathrm{E}[X] &= \sum_x x f_X(x) = 2.8 \\ \mathrm{E}[X^2] &= \sum_x x^2 f_X(x) = 8.5 \\ \mathrm{Var}(X) &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = 0.66 \end{aligned}$$

Similarly, $\mathrm{E}[Y] = 51$, $\mathrm{E}[Y^2] = 2670$, and $\mathrm{Var}(Y) = 69$.

$$\begin{aligned} \mathrm{E}[XY] &= \sum_x \sum_y xyf(x,y) \\ &= 2(40)(.0) + \cdots + 4(60)(.1) = 140 \\ \mathrm{Cov}(X,Y) &= \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] = -2.8 \\ \rho &= \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} = -0.415. \end{aligned}$$

Cts Example: Suppose $f(x,y) = 10x^2y$, $0 \le y \le x \le 1$.

$$\begin{aligned} f_X(x) &= \int_0^x 10x^2y\,dy \;=\; 5x^4, \;\; 0 \le x \le 1 \\ \mathrm{E}[X] &= \int_0^1 5x^5\,dx \;=\; 5/6 \\ \mathrm{E}[X^2] &= \int_0^1 5x^6\,dx \;=\; 5/7 \\ \mathrm{Var}(X) &= \mathrm{E}[X^2] - (\mathrm{E}[X])^2 \;=\; 0.01984 \end{aligned}$$

Similarly,

$$
\begin{aligned}
f_Y(y) &= \int_y^1 10x^2y\,dx \;=\; \frac{10}{3}y(1-y^3), \;\; 0 \le y \le 1 \\
\mathrm{E}[Y] &= 5/9, \quad \mathrm{Var}(Y) \;=\; 0.04850 \\
\mathrm{E}[XY] &= \int_0^1 \int_0^x 10x^3y^2\,dy\,dx \;=\; 10/21 \\
\mathrm{Cov}(X,Y) &= \mathrm{E}[XY] - \mathrm{E}[X]\mathrm{E}[Y] \;=\; 0.01323 \\
\rho &= \frac{\mathrm{Cov}(X,Y)}{\sqrt{\mathrm{Var}(X)\mathrm{Var}(Y)}} \;=\; 0.4265
\end{aligned}
$$

Theorems Involving Covariance

Theorem: $\text{Var}(X+Y) = \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y)$, *whether or not* X and Y are indep.

Remark: If X, Y are indep, the Cov term goes away.

Proof: By the work we did on a previous proof,

$$\begin{aligned}\text{Var}(X+Y) &= \text{E}[X^2] - (\text{E}[X])^2 + \text{E}[Y^2] - (\text{E}[Y])^2 \\ &\qquad +2(\text{E}[XY] - \text{E}[X]\text{E}[Y]) \\ &= \text{Var}(X) + \text{Var}(Y) + 2\text{Cov}(X,Y).\end{aligned}$$

Theorem:

$$\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) \;=\; \sum_{i=1}^{n} \mathrm{Var}(X_i) + 2\sum\sum\nolimits_{i<j} \mathrm{Cov}(X_i, X_j).$$

Proof: Induction.

Remark: If all X_i's are *indep*, then

$$\mathrm{Var}\Big(\sum_{i=1}^{n} X_i\Big) \;=\; \sum_{i=1}^{n} \mathrm{Var}(X_i).$$

Theorem: $\text{Cov}(aX, bY) = ab\text{Cov}(X, Y)$.

Proof:

$$\begin{aligned}\text{Cov}(aX, bY) &= \text{E}[aX \cdot bY] - \text{E}[aX]\text{E}[bY] \\ &= ab\text{E}[XY] - ab\text{E}[X]\text{E}[Y] \\ &= ab\text{Cov}(X, Y).\end{aligned}$$

Theorem:

$$\text{Var}(\sum_{i=1}^{n} a_i X_i) = \sum_{i=1}^{n} a_i^2 \text{Var}(X_i) + 2\sum\sum\nolimits_{i<j} a_i a_j \text{Cov}(X_i, X_j).$$

Proof: Put above two results together.

Example: $\text{Var}(X-Y) = \text{Var}(X) + \text{Var}(Y) - 2\text{Cov}(X,Y)$.

Example:

$$\begin{aligned}
&\text{Var}(X-2Y+3Z) \\
&\quad = \text{Var}(X) + 4\text{Var}(Y) + 9\text{Var}(Z) \\
&\qquad\qquad -4\text{Cov}(X,Y) + 6\text{Cov}(X,Z) - 12\text{Cov}(Y,Z).
\end{aligned}$$

Outline

1. Intro / Definitions
2. Discrete Random Variables
3. Continuous Random Variables
4. Cumulative Distribution Functions
5. Great Expectations
6. Functions of a Random Variable
7. Bivariate Random Variables
8. Conditional Distributions
9. Independent Random Variables
10. Conditional Expectation
11. Covariance and Correlation
12. Moment Generating Functions

Introduction

Recall that $\mathrm{E}[X^k]$ is the kth **moment** of X.

Definition: $M_X(t) \equiv \mathrm{E}[e^{tX}]$ is the **moment generating function** (mgf) of the RV X.

Remark: $M_X(t)$ is a function of t, *not* of X!

Example: $X \sim \mathrm{Bern}(p)$.

$$X = \begin{cases} 1 & \text{w.p. } p \\ 0 & \text{w.p. } q \end{cases}$$

Then

$$M_X(t) = \mathrm{E}[e^{tX}] = \sum_x e^{tx} f(x) = e^{t \cdot 1} p + e^{t \cdot 0} q = pe^t + q.$$

Example: $X \sim \mathrm{Exp}(\lambda)$.

$$f(x) = \begin{cases} \lambda e^{-\lambda x} & \text{if } x > 0 \\ 0 & \text{otherwise} \end{cases}$$

Then

$$\begin{aligned} M_X(t) &= \mathrm{E}[e^{tX}] \\ &= \int_{\mathbb{R}} e^{tx} f(x)\,dx \\ &= \lambda \int_0^{\infty} e^{(t-\lambda)x}\,dx \\ &= \frac{\lambda}{\lambda - t} \quad \text{if } \lambda > t. \end{aligned}$$

Big Theorem: Under certain technical conditions,

$$\mathrm{E}[X^k] \;=\; \frac{d^k}{dt^k} M_X(t)\bigg|_{t=0}, \quad k = 1, 2, \ldots .$$

Thus, you can *generate* the moments of X from the mgf. (Sometimes, it's easier to get moments this way than directly.)

"Proof:"

$$M_X(t) = \mathrm{E}[e^{tX}] = \mathrm{E}\left[\sum_{k=0}^{\infty}\frac{(tX)^k}{k!}\right] = \sum_{k=0}^{\infty}\mathrm{E}\left[\frac{(tX)^k}{k!}\right]$$
$$= 1 + t\mathrm{E}[X] + \frac{t^2\mathrm{E}[X^2]}{2} + \cdots$$

This implies

$$\frac{d}{dt}M_X(t) = \mathrm{E}[X] + t\mathrm{E}[X^2] + \cdots$$

and so

$$\frac{d}{dt}M_X(t)\bigg|_{t=0} = \mathrm{E}[X].$$

Same deal for higher-order moments.

Example: $X \sim \mathrm{Exp}(\lambda)$. Then $M_X(t) = \frac{\lambda}{\lambda - t}$ for $\lambda > t$. So

$$\mathrm{E}[X] = \frac{d}{dt} M_X(t)\Big|_{t=0} = \frac{\lambda}{(\lambda - t)^2}\Big|_{t=0} = 1/\lambda.$$

Further,

$$\mathrm{E}[X^2] = \frac{d^2}{dt^2} M_X(t)\Big|_{t=0} = \frac{2\lambda}{(\lambda - t)^3}\Big|_{t=0} = 2/\lambda^2.$$

Thus,

$$\mathrm{Var}(X) = \mathrm{E}[X^2] - (\mathrm{E}[X])^2 = \frac{2}{\lambda^2} - \frac{1}{\lambda^2} = 1/\lambda^2.$$

Other Applications

You can do lots of nice things with mgf's...

Find the mgf of a linear function of X.

Find the mgf of the sum of indep RV's.

Identify distributions.

Derive cool properties of distributions.

Theorem: Suppose X has mgf $M_X(t)$ and let $Y = aX + b$. Then $M_Y(t) = e^{tb} M_X(at)$.

Proof:

$$\begin{aligned} M_Y(t) &= \mathrm{E}[e^{tY}] \;=\; \mathrm{E}[e^{t(aX+b)}] \;=\; e^{tb}\mathrm{E}[e^{(at)X}] \\ &= e^{tb} M_X(at). \end{aligned}$$

Example: Let $X \sim \mathrm{Exp}(\lambda)$ and $Y = 3X + 2$. Then

$$M_Y(t) \;=\; e^{2t} M_X(3t) \;=\; e^{2t} \frac{\lambda}{\lambda - 3t}, \quad \text{if } \lambda > 3t.$$

Theorem: Suppose $X_1, \ldots, X_n$ are *indep*. Let $Y = \sum_{i=1}^n X_i$. Then

$$M_Y(t) \; = \; \prod_{i=1}^n M_{X_i}(t).$$

Proof:

$$\begin{aligned} M_Y(t) &= \mathrm{E}[e^{tY}] \; = \; \mathrm{E}[e^{t \sum X_i}] \; = \; \mathrm{E}\Big[\prod_{i=1}^n e^{tX_i}\Big] \\ &= \prod_{i=1}^n \mathrm{E}[e^{tX_i}] \quad (X_i\text{'s indep}) \\ &= \prod_{i=1}^n M_{X_i}(t). \end{aligned}$$

Corollary: If $X_1, \ldots, X_n$ are i.i.d. and $Y = \sum_{i=1}^n X_i$, then

$$M_Y(t) \;=\; [M_{X_1}(t)]^n.$$

Example: Suppose $X_1, \ldots, X_n \overset{\text{iid}}{\sim} \text{Bern}(p)$. Then by a previous example,

$$M_Y(t) \;=\; [M_{X_1}(t)]^n \;=\; (pe^t + q)^n.$$

So what use is a result like this?

Theorem: If X and Y have the same mgf, then they have the *same distribution* (at least in this course)!

Proof: Too hard for this course.

Example: The sum Y of n i.i.d. Bern(p) RV's is the same as a Binomial(n, p) RV.

By the previous example, $M_Y(t) = (pe^t + q)^n$. So all we need to show is that the mgf of the matches this.

Meanwhile, let $Z \sim \mathrm{Bin}(n,p)$.

$$
\begin{aligned}
M_Z(t) &= \mathrm{E}[e^{tZ}] \;=\; \sum_z e^{tz} P(Z=z) \\
&= \sum_{z=0}^{n} e^{tz} \binom{n}{z} p^z q^{n-z} \\
&= \sum_{z=0}^{n} \binom{n}{z} (pe^t)^z q^{n-z} \\
&= (pe^t + q)^n \quad \text{(by the binomial theorem)},
\end{aligned}
$$

and this matches the mgf of Y from the last pg.

Example: You can identify a distribution by its mgf.

$$M_X(t) \; = \; \left(\frac{3}{4}e^t + \frac{1}{4}\right)^{15}$$

implies that $X \sim \text{Bin}(15, 0.75)$.

Example:

$$M_Y(t) \; = \; e^{-2t}\left(\frac{3}{4}e^{3t} + \frac{1}{4}\right)^{15}$$

implies that Y has the same distribution as $3X - 2$, where $X \sim \text{Bin}(15, 0.75)$.

Theorem (Additive property of Binomials): If $X_1, \ldots, X_k$ are indep, with $X_i \sim \mathrm{Bin}(n_i, p)$ (where p is the same for all X_i's), then

$$Y \equiv \sum_{i=1}^{k} X_i \sim \mathrm{Bin}\Big(\sum_{i=1}^{k} n_i, p\Big).$$

Proof:

$$\begin{aligned} M_Y(t) &= \prod_{i=1}^{k} M_{X_i}(t) \quad \text{(mgf of indep sum)} \\ &= \prod_{i=1}^{k} (pe^t + q)^{n_i} \quad \text{(Binomial}(n_i, p)\text{ mgf)} \\ &= (pe^t + q)^{\sum_{i=1}^{k} n_i}. \end{aligned}$$

This is the mgf of the Bin$(\sum_{i=1}^{k} n_i, p)$, so we're done.