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

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

The mean gives an indication of a RV's *central tendency*.

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

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

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

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

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

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

Law of the Unconscious Statistician

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

$$\mathsf{E}[g(X)] \equiv \begin{cases} \sum_x g(x)f(x) & \text{if } X \text{ is discrete} \\ \int_{\Re} g(x)f(x)\,dx & \text{if } X \text{ is cts} \end{cases}$$

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

$\mathsf{E}[\sin X] = \int_{\Re} (\sin x) f(x)\,dx$

Just a moment please...

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

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

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

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

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

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

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

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

Proof (just do cts case):

$$\begin{aligned}\mathsf{E}[ag(X)+b] &= \int_{\Re}(ag(x)+b)f(x)\,dx \\ &= a\int_{\Re} g(x)f(x)\,dx + b\int_{\Re} f(x)\,dx \\ &= a\mathsf{E}[g(X)]+b.\end{aligned}$$

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

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

Theorem (easier way to calculate variance):

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

Proof:

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

Example: $X \sim$ Bern(p).

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

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

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

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

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

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

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

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

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

Proof:

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

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 $Y = g(X) = 4X + 5$. Then

$\text{E}[Y] = \text{E}[4X + 5] = 4\text{E}[X] + 5 = 6.2$
$\text{Var}(Y) = \text{Var}(4X + 5) = 16\text{Var}(X) = 3.36$.

Chebychev's Inequality

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

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

Proof: See text.

Remarks:

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

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

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 $X \sim \mathsf{U}(0,1)$. $f(x) = 1$, $0 < x < 1$.

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

Then Chebychev implies

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

In particular, for $\epsilon = 1/4$,

$$\Pr(|X - \frac{1}{2}| \geq \frac{1}{4}) \ \leq \ \frac{4}{3} \ \ \text{(TERRIBLE bound!)}.$$

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

$$
\begin{aligned}
&\Pr(|X - \tfrac{1}{2}| \geq \tfrac{1}{4}) \\
&= 1 - \Pr(|X - \tfrac{1}{2}| < \tfrac{1}{4}) \\
&= 1 - \Pr(-\tfrac{1}{4} < X - \tfrac{1}{2} < \tfrac{1}{4}) \\
&= 1 - \Pr(\tfrac{1}{4} < X < \tfrac{3}{4}) \\
&= 1 - \int_{1/4}^{3/4} f(x)\,dx \\
&= 1 - \tfrac{1}{2} = 1/2.
\end{aligned}
$$