5.28 Bivariate Normal and Friends

Bivariate Normal Distrn

Bivariate Normal Example

Lognormal Distrn

Simulating Normal RV's

(X, Y) has the **Bivariate Normal Distrn** if it has p.d.f.

$$f(x,y) = C \exp\left\{ \frac{-\left[z_X^2(x) - 2\rho z_X(x) z_Y(y) + z_Y^2(y)\right]}{2(1-\rho^2)} \right\}$$

where

$$\rho \equiv \text{Corr}(X,Y), \quad C \equiv \frac{1}{2\pi\sigma_X\sigma_Y\sqrt{1-\rho^2}},$$

$$z_X(x) \equiv \frac{x-\mu_X}{\sigma_X} \text{ and } z_Y(y) \equiv \frac{y-\mu_Y}{\sigma_Y}.$$

Pretty nasty joint p.d.f., eh?

In fact, $X \sim \text{Nor}(\mu_X, \sigma_X^2)$ and $Y \sim \text{Nor}(\mu_Y, \sigma_Y^2)$.

Example: (X, Y) could be a person's (height,weight). The two quantities are marginally normal, but positively correlated.

If you want to calculate bivariate normal probabilities, you'll need to evaluate quantities like

$$\Pr(a < X < b, c < Y < d) = \int_c^d \int_a^b f(x,y)\,dx\,dy,$$

which will probably require numerical integration techniques.

Fun Fact (which will come up later when we discuss regression): The conditional distribution of Y given that $X = x$ is also normal. In particular,

$$Y|X = x \sim \text{Nor}(\mu_Y + \rho(\sigma_Y/\sigma_X)(x - \mu_X), \sigma_Y^2(1 - \rho^2)).$$

Information about X helps to update the distribution of Y.

Example: Consider students at a university. Let X be their combined SAT scores (Math and Verbal), and Y their freshman GPA (out of 4). Suppose a study reveals that

$$\mu_X = 1300, \quad \mu_Y = 2.3,$$

$$\sigma_X^2 = 6400, \quad \sigma_Y^2 = 0.25, \quad \rho = 0.6.$$

Find $\Pr(Y \geq 2 | X = 900)$.

First,

$$\begin{aligned} & \mathsf{E}[Y|X = 900] \\ & \quad = \mu_Y + \rho(\sigma_Y/\sigma_X)(x - \mu_X) \\ & \quad = 2.3 + \rho(\sqrt{0.25/6400})(900 - 1300) = 0.8, \end{aligned}$$

indicating that the expected GPA of a kid with 900 SAT's will be 0.8.

Second,

$$\mathsf{Var}(Y|X = 900) = \sigma_Y^2(1 - \rho^2) = 0.16.$$

Thus,

$$Y|X = 900 \ \sim \ \text{Nor}(0.8, 0.16).$$

Now we can calculate

$$\begin{aligned} \Pr(Y \geq 2|X = 900) &= \Pr\left(Z \geq \frac{2 - 0.8}{\sqrt{0.16}}\right) \\ &= 1 - \Phi(3) = 0.0013. \end{aligned}$$

This guy doesn't have much chance of having a good GPA.

Lognormal Distrn

Definition: If $Y \sim \text{Nor}(\mu_Y, \sigma_Y^2)$, then $X \equiv e^Y$ has the **lognormal distrn** with parameters (μ_Y, σ_Y^2).

Turns Out: The p.d.f. of the lognormal is

$$f(x) = \frac{1}{x\sigma_Y\sqrt{2\pi}} \exp\left\{-\frac{1}{2\sigma_Y^2}[\ell\text{n}(x) - \mu_Y]^2\right\}, \quad x > 0.$$

Further,

$$\mathsf{E}[X] = \exp\left\{\mu_Y + \frac{\sigma_Y^2}{2}\right\}$$

$$\mathsf{Var}(X) = \exp\{2\mu_Y + \sigma_Y^2\}\Big(\exp\{\sigma_Y^2\} - 1\Big)$$

The lognormal distrn has lots of other nice properties.

See any good prob/stats book for details.

Example: Suppose $Y \sim \text{Nor}(10, 4)$ and let $X = e^Y$. Then

$$\Pr(Y \leq \ell\text{n}(1000)) = \Pr\left(Z \leq \frac{\ell\text{n}(1000) - 10}{2}\right) = \Phi(-1.55) = 0.061.$$

Simulating Normal RV's

How would you generate a normal RV's when conducting computer simulation experiments?

Theorem (Box and Müller): If $U_1, U_2 \overset{\text{iid}}{\sim} \text{U}(0,1)$, then

$$\begin{aligned} Z_1 &= \sqrt{-2\ell\text{n}(U_1)}\cos(2\pi U_2), \\ Z_2 &= \sqrt{-2\ell\text{n}(U_1)}\sin(2\pi U_2) \end{aligned}$$

are iid Nor(0,1).

Remarks: (1) Proof: Not here.

(2) Many other ways to generate Nor(0,1)'s, but this is the easiest.

(3) Cosine and Sine must be calculated in *radians*, not degrees.

(4) To get $X \sim$ Nor(μ, σ^2), just take $X = \mu + \sigma Z$.