Lecture 37: Dec 6

Last time

- Expectation

Today

- Course Evaluations (13/48)
- Transformations

Bivariate Transformations of Continuous RVs Suppose (X, Y) is continuous and the joint transformation

$$u = g_1(x, y), \quad v = g_2(x, y)$$

is one-to-one and differentiable. Define the inverse mapping

$$x = h_1(u, v), \quad y = h_2(u, v)$$

Then

$$f_{UV}(u, v) = f_{XY}\left(h_1(u, v), h_2(u, v)\right) ||J(u, v)||$$

where $J(u, v)$ is the Jacobian of the transformation $(x, y) \to (u, v)$ given by

$$J(u, v) = \frac{\partial(x, y)}{\partial(u, v)} = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix}$$

Example: Rotation of a bivariate normal vector Let $X \sim N(0, 1)$, $Y \sim N(0, 1)$, independent. Define the rotation

$$U = X\cos\theta - Y\sin\theta$$
$$V = X\sin\theta + Y\cos\theta$$

for fixed θ. Then $U \sim N(0, 1)$, $V \sim N(0, 1)$, independent.

Proof:
The range of (X, Y) is $\mathbb{R}^2$. The range of (U, V) is $\mathbb{R}^2$. Need the inverse transformation

$$X = U\cos\theta + V\sin\theta$$
$$Y = -U\sin\theta + V\cos\theta$$

with Jacobian

$$J(u, v) = \begin{bmatrix} \frac{\partial x}{\partial u} & \frac{\partial x}{\partial v} \\ \frac{\partial y}{\partial u} & \frac{\partial y}{\partial v} \end{bmatrix} = \begin{bmatrix} \cos\theta & \sin\theta \\ -\sin\theta & \cos\theta \end{bmatrix}$$

and its determinant is

$$|J(u, v)| = 1.$$

The joint pdf of (X, Y) is

$$f_{XY}(x, y) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-y^2/2} = \frac{1}{2\pi} e^{-(x^2+y^2)/2}$$

The joint pdf of (U, V) is

$$\begin{aligned} f_{UV}(u, v) &= \frac{1}{2\pi} e^{-\left[(u\cos\theta + v\sin\theta)^2 + (-u\sin\theta + v\cos\theta)^2\right]/2} \cdot |1| \\ &= \frac{1}{2\pi} e^{-(u^2+v^2)/2} = \frac{1}{\sqrt{2\pi}} e^{-u^2/2} \cdot \frac{1}{\sqrt{2\pi}} e^{-v^2/2} \end{aligned}$$

so $U \sim N(0,1)$, $V \sim N(0,1)$, and U and V are independent.

Functions of independent random variables (Theorem 4.3.5) Let X and Y be independent rvs. Let $g : \mathbb{R} \to \mathbb{R}$ and $h : \mathbb{R} \to \mathbb{R}$ be functions. Then the random variables $U = g(X)$ and $V = h(Y)$ are independent.

Sum of two independent rvs Suppose X and Y are independent. What is the distribution of $Z = X + Y$? In general:

$$F_Z(z) = \Pr(X + Y \leqslant z) = \Pr(\{(x, y) \text{ such that } x + y \leqslant z\})$$

Various approaches:

- bivariate transformation method (continuous and discrete)
- Discrete convolution

$$f_Z(z) = \sum_{x+y=z} f_X(x) f_Y(y) = \sum_x f_X(x) f_Y(z - x)$$

- Continuous convolution (Section 5.2)
- MGF method (continuous and discrete)

Example (Sum of two independent Poissons) Define X, Y to be two independent random variables having Poisson distributions with parameters λ_i, $i = 1, 2$. Then:

$$f_{X,Y}(x, y) = \frac{e^{-\lambda_1}\lambda_1^x}{x!} \frac{e^{-\lambda_2}\lambda_2^y}{y!}, x, y = 0, 1, 2, \ldots$$

The distribution of $S = X + Y$ is

$$\begin{aligned} f_S(s) &= \sum_{x=0}^{s} \frac{e^{-\lambda_1}\lambda_1^x}{x!} \frac{e^{-\lambda_2}\lambda_2^{s-x}}{(s-x)!} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{s!} \sum_{x=0}^{s} \binom{s}{x} \lambda_1^x \lambda_2^{s-x} \\ &= \frac{e^{-(\lambda_1+\lambda_2)}}{s!} (\lambda_1 + \lambda_2)^s \end{aligned}$$

Again, S is Poisson with parameter $\lambda = \lambda_1 + \lambda_2$.

Moment generating function (Theorem 4.2.12) Let X and Y be independent rvs with mgfs $M_X(\cdot)$ and $M_Y(\cdot)$, respectively. Then the mgf of $Z = X + Y$ is

$$M_Z(t) = M_X(t)M_Y(t)$$

Proof:

$$\begin{aligned} M_Z(t) &= E\exp(Zt) &&= E\{\exp[(X+Y)t]\} \\ &= E[\exp(Xt)\exp(Yt)] &&= E[\exp(Xt)] \cdot E[\exp(Yt)] \\ &= M_X(t)M_Y(t) \end{aligned}$$

Corollary: If X and Y are independent and $Z = X - Y$,

$$M_Z(t) = M_X(t)M_Y(-t)$$

Example (sum of two independent Poissons) Suppose $X \sim Poisson(\lambda_X)$ and $Y \sim Poisson(\lambda_Y)$ and put $Z = X + Y$. Then, $Z \sim Poisson(\lambda_X + \lambda_Y)$. *Proof:*

$$\begin{aligned} M_Z(t) &= \exp\left[\lambda_X(e^t - 1)\right]\exp\left[\lambda_Y(e^t - 1)\right] \\ &= \exp\left[(\lambda_X + \lambda_Y)(e^t - 1)\right] \end{aligned}$$

Example (sum of two independent normals) Suppose $X \sim N(\mu_x, \sigma_x^2)$ and $Y \sim N(\mu_y, \sigma_y^2)$ and X and Y are independent and $Z = X + Y$. Then

$$Z \sim N(\mu_x + \mu_y, \sigma_x^2 + \sigma_y^2)$$

Proof:

$$\begin{aligned} M_Z(t) &= \exp\left(\mu_x t + \frac{1}{2}\sigma_x^2 t^2\right)\exp\left(\mu_y t + \frac{1}{2}\sigma_y^2 t^2\right) \\ &= \exp\left[(\mu_x + \mu_y)t + \frac{1}{2}(\sigma_x^2 + \sigma_y^2)t^2\right] \end{aligned}$$

Example (sum of two independent gammas) Suppose $X \sim \Gamma(\alpha_x, \beta)$ and independently $Y \sim \Gamma(\alpha_y, \beta)$. Let $Z = X + Y$. Then $Z \sim \Gamma((\alpha_x + \alpha_y), \beta)$.
Proof:

$$\begin{aligned} M_Z(t) &= \left(\frac{1}{1-\beta t}\right)^{\alpha_x}\left(\frac{1}{1-\beta t}\right)^{\alpha_y} \\ &= \left(\frac{1}{1-\beta t}\right)^{\alpha_x + \alpha_y} \end{aligned}$$

Remember that

- If $\alpha = 1$ we have an exponential with parameter β.
- If $\alpha = n/2$ and $\beta = 2$, we have a $\chi^2(n)$ (with n d.f.). The above result states that $\chi^2(n_1) + \chi^2(n_2) = \chi^2(n_1 + n_2)$.

Covariance and Correlation Let X and Y be two random variables with respective means μ_X, μ_Y and variances $\sigma_X^2 > 0$ and $\sigma_Y^2 > 0$, all assumed to exist.

- The *covariance* of X and Y is

$$Cov(X, Y) = E\left[(X - \mu_X)(Y - \mu_Y)\right] = \sigma_{XY}$$

- The *correlation* between X and Y is

$$Cor(X, Y) = \frac{Cov(X, Y)}{\sqrt{Var(X)Var(Y)}}$$

also written as

$$\rho_{XY} = \frac{\sigma_{XY}}{\sigma_X \sigma_Y} = E\left[\left(\frac{X - \mu_X}{\sigma_X}\right)\left(\frac{Y - \mu_Y}{\sigma_Y}\right)\right]$$

Properties Let c be a constant:

1. $Cov(X, X) = Var(X)$, $\quad Cor(X, X) = 1$
2. $Cov(X, Y) = Cov(Y, X)$, $\quad Cor(X, Y) = Cor(Y, X)$
3. $Cov(X, c) = 0$, $\quad Cor(X, c) = 0$
4. $Cov(X, Y) = E(XY) - E(X)E(Y)$

5. Let $X_c = X - \mu_X$, $Y_c = Y - \mu_Y$. Then

$$\begin{aligned} Cov(X, Y) &= Cov(X_c, Y_c) = E(X_c Y_c) \\ Cor(X, Y) &= Cor(X_c, Y_c) \end{aligned}$$

6. Let $\tilde{X} = (X - \mu_X)/\sigma_X$, $\tilde{Y} = (Y - \mu_Y)/\sigma_Y$. Then,

$$Cor(X, Y) = Cor(\tilde{X}, \tilde{Y}) = Cov(\tilde{X}, \tilde{Y}) = E(\tilde{X}\tilde{Y})$$

Independent vs. Uncorrelated

- X and Y are called *uncorrelated* iff

$$Cov(X, Y) = 0 \quad \text{or equivalently} \quad \rho_{XY} = 0$$

- If X and Y are independent and $Cov(X, Y)$ exists, then $Cov(X, Y) = 0$.
- If X and Y are uncorrelated, this does **not** imply that they are independent.

Example $X \sim U[-1, 1]$, $Y = X^2$. Then $Cov(X, Y) = 0$ but X, Y are not independent.

Correlation coefficient For any random variables X and Y,

1. $-1 \leqslant \rho_{XY} \leqslant 1$
2. $|\rho_{XY}| = 1$ if and only if $\exists a \neq 0$ and b such that

$$\Pr(Y = aX + b) = 1.$$

   if $\rho_{XY} = 1$ then $a > 0$, and if $\rho_{XY} = -1$, then $a < 0$.

proof:
Let $\tilde{X} = (X - \mu_X)/\sigma_X$, $\tilde{Y} = (Y - \mu_Y)/\sigma_Y$. Then $Cor(X, Y) = E(\tilde{X}\tilde{Y})$,

1.
$$\begin{aligned} 0 \leqslant E(\tilde{X} - \tilde{Y})^2 = 1 + 1 - 2E(\tilde{X}\tilde{Y}) &\Rightarrow \quad E(\tilde{X}\tilde{Y}) \leqslant 1 \\ 0 \leqslant E(\tilde{X} + \tilde{Y})^2 = 1 + 1 + 2E(\tilde{X}\tilde{Y}) &\Rightarrow \quad -1 \leqslant E(\tilde{X}\tilde{Y}) \end{aligned}$$

2.
$$\begin{aligned} \rho_{XY} = 1 &\iff \Pr(\tilde{Y} = \tilde{X}) = 1 \qquad \Rightarrow \quad a > 0 \\ \rho_{XY} = -1 &\iff \Pr(\tilde{Y} = -\tilde{X}) = 1 \quad \Rightarrow \quad a < 0 \end{aligned}$$

Random Samples

Definition The random variables $X_1, \dots, X_n$ are called a *random sample of size n from the population $f(x)$* if $X_1, \dots, X_n$ are mutually independent and identically distributed (iid) random variables with the same pdf or pmf $f(x)$.

If $X_1, \dots, X_n$ are iid, then their joint pdf or pmf is

$$f(x_1, \dots, x_n) = f(x_1)f(x_2)\dots f(x_n) = \prod_{j=1}^{n} f(x_j)$$

Statistics Let $X_1, \dots, X_n$ be a random sample and let $T(x_1, \dots, x_n)$ be a function defined on $\mathbb{R}^n$. Then the random variable $Y = T(X_1, \dots, X_n)$ is called a *statistic*. The probability distribution of Y is called the *sampling distribution* of Y.

Note: T is only a function of $(x_1, \dots, x_n)$, no parameters.

Examples

$$\begin{aligned} \text{sample mean} \quad & \bar{X} = \tfrac{1}{n} \sum_{j=1}^{n} X_j \\ \text{sample variance} \quad & S^2 = \tfrac{1}{n-1} \sum_{j=1}^{n} (X_j - \bar{X})^2 \\ \text{sample standard deviation} \quad & S = \sqrt{S^2} \\ \text{minimum} \quad & X_{(1)} = \min_{1 \leqslant i \leqslant n} X_i \end{aligned}$$

Properties Let $x_1, \ldots, x_n$ be n numbers and define

$$\bar{x} = \frac{1}{n}\sum_{j=1}^{n} x_j, \quad s^2 = \frac{1}{n-1}\sum_{j=1}^{n}(x_j - \bar{x})^2$$

Then

$$\min_a \sum_{j=1}^{n}(x_j - a)^2 = \sum_{j=1}^{n}(x_j - \bar{x})^2$$

$$(n-1)s^2 = \sum_{j=1}^{n}(x_j - \bar{x})^2 = \sum_{j=1}^{n} x_j^2 - n\bar{x}^2$$

Residuals Lemma: Let $X_1, \ldots, X_n$ be a random sample from a population with mean μ and variance σ^2. Define the residuals $R_i = X_i - \bar{X}$. Then

$$E(R_i) = 0, \quad Var(R_i) = \tfrac{n-1}{n}\sigma^2$$
$$Cov(R_i, \bar{X}) = 0, \quad Cov(R_i, R_j) = -\sigma^2/n \text{ if } i \neq j$$

Theorem Let $X_1, \ldots, X_n$ be a random sample from a population with mgf $M_X(t)$. Then the mgf of the sample mean is

$$M_{\bar{X}}(t) = [M_X(t/n)]^n$$

Convergence

Convergence in Probability A sequence of random variables $X_1, \ldots, X_n$ *converges in probability* to a random variable X, denoted

$$X_n \xrightarrow{p} X$$

if for every $\epsilon > 0$,

$$\lim_{n\to\infty} \Pr(|X_n - X| < \epsilon) = 1$$

or equivalently

$$\lim_{n\to\infty} \Pr(|X_n - X| > \epsilon) = 0$$

In other words, X_n is more and more likely to be close to X, or less and less likely to be far from X.

Example Let $X_n = X + \epsilon_n$, where $\epsilon_n \sim N(0, 1/n)$ and X is an arbitrary random variable. Then, as $n \to \infty$,

$$X_n \xrightarrow{p} X$$

Weak law of large numbers (WLLN) Let $Y_1, \dots, Y_n$ be iid with common mean μ and variance σ^2. Then, as $n \to \infty$,

$$\bar{Y}_n = \frac{1}{n}\sum_{j=1} Y_j \xrightarrow{p} \mu$$

Proof:
The proof is quite simple, being a straightforward application of Chebychev's Inequality. We have, for every $\epsilon > 0$,

$$\Pr(|\bar{Y}_n - \mu| \geqslant \epsilon) = \Pr(|\bar{Y}_n - \mu|^2 \geqslant \epsilon^2) \leqslant \frac{E(\bar{Y} - \mu)^2}{\epsilon^2} = \frac{Var(\bar{Y})}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0 \text{ as } n \to \infty$$

Convergence in Distribution A sequence of random variables $X_1, \dots, X_n$ *converges in distribution* to a random variable X, denoted

$$X_N \xrightarrow{d} X$$

if

$$\lim_{n\to\infty} F_{X_n}(x) = F_X(x)$$

This is also called *convergence in law* or *weak convergence*. In other words, the distribution of X_n is closer and closer to the distribution of X.

Relation between "in distribution" and "in probability" Theorem:

1. Convergence in probability implies convergence in distribution:

$$X_n \xrightarrow{p} X \Rightarrow X_n \xrightarrow{d} X$$

2. Suppose $X_n \xrightarrow{d} X$ where X has a degenerate distribution, i.e. $\Pr\{X = a\} = 1$ for some $a \in \mathbb{R}$. Then,

$$X_n \xrightarrow{d} a \Rightarrow X_n \xrightarrow{p} a$$

Convergence in Distribution via Convergence of Mgfs Theorem: Suppose the mgf $M_n(t)$ of Y_n exists for $|t| < h$, and the mgf $M(t)$ of Y exists for $|t| < h_1 < h$. Then,

$$Y_n \xrightarrow{d} Y \iff \lim_{n\to\infty} M_n(t) = M(t), \quad |t| < h_1$$

Example Let $X_\lambda \sim Poisson(\lambda)$. Then, as $\lambda \to \infty$,

$$\frac{X_\lambda - \lambda}{\lambda} \xrightarrow{p} 0$$
$$\frac{X_\lambda - \lambda}{\sqrt{\lambda}} \xrightarrow{d} N(0,1)$$

Central Limit Theorem   Let $X_1, X_2, \ldots, X_n$ be a sequence of iid random variables whose mgfs exist in a neighborhood of 0 (that is, $M_{X_i}(t)$ exists for $|t| < h$, for some positive $h > 0$). Let $EX_i = \mu$ and $Var(X_i) = \sigma^2 > 0$. (Both μ and σ^2 are finite since the mgf exists) Define $\bar{X}_n = \frac{1}{n}\sum_{i=1}^{n} X_i$. Let $G_n(x)$ denote the cdf of $\sqrt{n}(\bar{X}_n - \mu)/\sigma$. Then, for any x, $-\infty < x < \infty$,

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} e^{-y^2/2} dy;$$

that is, $\sqrt{n}(\bar{X}_n - \mu)/\sigma$ has a limiting standard normal distribution, in other words, $\sqrt{n}(\bar{X}_n - \mu)/\sigma \xrightarrow{d} N(0,1)$

Proof:
Define $Y_i = (X_i - \mu)/\sigma$, and let $M_Y(t)$ denote the common mgf of Y_is, which exists for $|t| < \sigma h$ and $M_Y(t) = M_{\frac{1}{\sigma}X_i - \mu/\sigma}(t) = e^{-\frac{\mu}{\sigma}t} M_X(\frac{t}{\sigma})$. Since

$$\frac{\sqrt{n}(\bar{X}_n)}{\sigma} = \frac{1}{\sqrt{n}} \sum_{i=1}^{n} Y_i,$$

we have,

$$\begin{aligned}
M_{\sqrt{n}(\bar{X}_n - \mu)/\sigma}(t) &= M_{\sum_{i=1}^{n} Y_i/\sqrt{n}}(t) \\
&= M_{\sum_{i=1}^{n} Y_i}(t/\sqrt{n}) \\
&= \left[M_Y(t/\sqrt{n})\right]^n.
\end{aligned}$$

We now expand $M_Y(t/\sqrt{n})$ in a Taylor series (power series) around 0.

$$M_Y(\frac{t}{\sqrt{n}}) = \sum_{k=0}^{\infty} M_Y^{(k)}(0) \frac{(t/\sqrt{n})^k}{k!},$$

where $M_Y^{(k)}(0) = (d^k/dt^k) M_Y(t)|_{t=0}$. Since the mgfs exist for $|t| < h$, the power series expansion is valid if $t < \sqrt{n}\sigma h$.

Using the facts that $M_Y^{(0)} = 1$, $M_Y^{(1)} = 0$, and $M_Y^{(2)} = 1$ (by construction, the mean and variance of Y are 0 and 1), we have

$$M_Y(\frac{t}{\sqrt{n}}) = 1 + \frac{(t/\sqrt{n})^2}{2!} + R_Y(\frac{t}{\sqrt{n}}),$$

where R_Y is the remainder term in the Taylor expansion such that

$$\lim_{n\to\infty} \frac{R_Y(t/\sqrt{n})}{(t/\sqrt{n})^2} = 0.$$

Therefore, for any fixed t, we can write

$$\begin{aligned}
\lim_{n\to\infty} \left[M_Y(\frac{t}{\sqrt{n}})\right]^n &= \lim_{n\to\infty} \left[1 + \frac{(t/\sqrt{n})^2}{2!} + R_Y(\frac{t}{\sqrt{n}})\right]^n \\
&= \lim_{n\to\infty} \left[1 + \frac{1}{n}\left(\frac{t^2}{2} + nR_Y(\frac{t}{\sqrt{n}})\right)\right]^n \\
&= e^{t^2/2}
\end{aligned}$$