ECON 626: Empirical Microeconomics
Department of Economics
University of Maryland
Fall 2019

Random variable review notes

1 Basics

For any constants a and b and random variables X and Y:

$$E[aX+b] = aE[X] + b$$
$$Var(aX+b) = a^2 Var(X)$$
$$E[X+Y] = E[X] + E[Y]$$
$$Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$$

Thus, for any **independent** random variables X and Y:

$$Var(X+Y) = Var(X) + Var(Y)$$

2 Normal distribution

Let $X \sim \mathcal{N}(\mu, \sigma^2)$. The probability density function (PDF) of X is given as:

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \qquad \forall x : -\infty \leq x \leq \infty$$

The cumulative distribution function (CDF) of X, $F_X(x) = \int_{-\infty}^{x} f_X(x)$, is denoted by $\Phi(\frac{x-\mu}{\sigma})$, where $\Phi(\cdot)$ is the CDF of the standard normal (mean zero, variance 1).

For any constants a and b and **independent**, **normally distributed** random variables X and Y, let $X \sim \mathcal{N}(\mu_X, \sigma_X^2)$ and let $Y \sim \mathcal{N}(\mu_Y, \sigma_Y^2)$. (The parameters of the normal distribution, specified this way, are the mean and the variance.) Then:

$$aX + b \sim \mathcal{N}(a\mu_X + b, a^2\sigma_X^2)$$
$$X + Y \sim \mathcal{N}(\mu_X + \mu_Y, \sigma_X^2 + \sigma_Y^2)$$

Thus, if we contruct $Z = \frac{X - \mu_X}{\sigma_X}$, the new random variable Z will be distributed according to the standard normal, $\mathcal{N}(0,1)$.

3 Uniform distribution

Let $X \sim \mathcal{U}(a, b)$. The probability density function (PDF) of X is given as:

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

For any constants a, b, c, and d and **uniform** random variable X, let $X \sim \mathcal{U}(a, b)$. (The parameters of the uniform distribution, specified this way, are the lower and upper limits of support.) Then:

$$\begin{aligned} cX &\sim \mathcal{U}(ca, cb) \\ X + d &\sim \mathcal{U}(a + d, b + d) \\ cX + d &\sim \mathcal{U}(ca + d, cb + d) \end{aligned}$$

4 Transformation

Recall that the cumulative distribution function (CDF) F_Y maps each number y on the real line to the probability of Y having a value no greater than y; that probability is a number between zero and one (inclusive). The inverse of F_Y, a function $F_Y^{-1}(p)$ thus maps a probability p $(0 \leq p \leq 1)$ back to the associated percentile of the Y distribution.

There is a particularly useful consequence of this for the purposes of simulating data generating processes. Suppose we have access to a uniform random variable, $X \sim \mathcal{U}(0, 1)$. Suppose we would like to have a random variable that is instead distributed according to CDF $F_Y(y)$. If we can analytically (or numerically) construct the inverse CDF function $F_Y^{-1}(\cdot)$, we only need to apply it to the uniform random variable X, which may be thought of as draws of percentiles: the new random variable constructed by the transformation $F_Y^{-1}(X)$ now has CDF $F_Y(y)$.

For further discussion, see any of:

- Casella and Berger, *Statistical Inference*, 2nd edition, Theorem 2.1.10 ("Probability integral transformation")
- Hogg and Tanis, *Probability and Statistical Inference*, 6th edition, Theorem 4.5-1, in the section "Distributions of Functions of a Random Variable."
- Cameron and Trivedi, *Microeconometrics*, 8th edition, Section 12.8.2, discussing the "inverse transformation" method as part of Section 12.8, "Methods of Drawing Random Variates."