Analysis of Economics Data
Appendix B: Some Essentials of Probability

© A. Colin Cameron
Univ. of Calif. Davis

November 2022

APPENDIX B: Some Essentials of Probability

- Appendix B considers properties of random variables.
- Probability theory for a single random variable
  - chapter 3 focused on a discrete random variable
    - e.g. coin toss
  - appendix also considers continuous random variables
    - e.g. normal and t distribution.
- Probability theory for the Sample Mean
  - more detail than Chapter 3.
- Probability theory for two related random variables
  - define formally concepts used in Chapter 6
    - conditional mean and conditional variance.
  - define the population analogs of sample covariance and sample correlation.

Outline

1. Probability Theory for a Single Random Variable
   1. Discrete Random Variables, Bernoulli Distribution
   2. Linear Transformation of a Random Variable
   3. Continuous Random Variable
   4. Standard Normal Distribution, Other Continuous Distributions
2. Probability Theory for the Sample Mean
   1. Statistical Independence
   2. Sums of Independent Random Variables
   3. Mean and Variance of the Sample Mean
   4. Law of Large Numbers, Central Limit Theorem
3. Probability Theory for Two Related Random Variables
   1. Joint Probability
   2. Conditional Distribution, Mean, Variance
   3. Covariance and Correlation

Appendix B.1: Single Random Variable

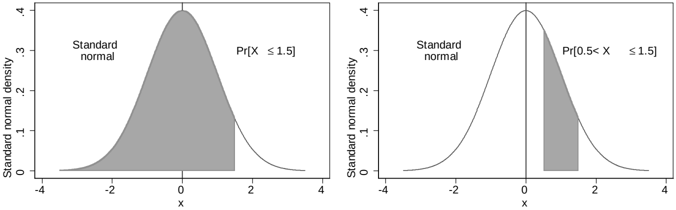
- We often use linear transformations of a random variable
  - key result: $Y = a + bX$ has mean $a + b\mathrm{E}[X]$ and variance $a^2\mathrm{Var}[X]$.
- Linear transformations of a random variable X
  - $\mathrm{E}[a] = a$ expected value of a constant is a constant
  - $\mathrm{E}[bX] = b\mathrm{E}[X]$ mean of b times X is b times the mean
  - $\mathrm{E}[a + bX] = a + b\mathrm{E}[X]$ combining
  - $\mathrm{Var}[bX] = b\mathrm{Var}[X]$ variance of b times X is b^2 times the variance
    - $\mathrm{E}[(bX - \mathrm{E}[bX])^2] = \mathrm{E}[\{b(X - \mathrm{E}[X])\}^2]$ $= b^2\mathrm{E}[(X - \mathrm{E}[X])^2] = b^2\mathrm{Var}[X]$
  - $\mathrm{Var}[a + bX] = b^2\mathrm{Var}[X]$.
- Standardization for $X \sim (\mu, \sigma^2)$
  - $\mathrm{E}[X - \mu] = \mathrm{E}[X] - \mu = \mu - \mu = 0$.
  - $\mathrm{Var}[\frac{X-\mu}{\sigma}] = \mathrm{Var}[\frac{X-\mu}{\sigma}] = (\frac{1}{\sigma})^2\mathrm{Var}[X - \mu] = (\frac{1}{\sigma})^2\sigma^2 = 1$.
  - so $\frac{X-\mu}{\sigma} \sim (0, 1)$
  - standardize by subtract mean and divide by standard deviation.

Continuous random variables use integration

- **Discrete** random variables take a limited number of values
  - $\Pr[X = x] > 0$ for all values that x may take
  - $\Pr[X \leq x] = \sum_{y \leq x} \Pr[X = y] =$ Sum of probabilities up to x
  - $\mathrm{E}[X] = \sum_x x \cdot \Pr[X = x]$ weight x by probabilities.
- **Continuous** random variables take uncountably infinite number of values, so individual probabilities do not exist.
  - $\Pr[X = x] = 0$ for all values that x may take
- **Continuous random variables replace summation with integration**
  - $\Pr[X \leq x] = \int_{-\infty}^{x} f(y)dy =$ Area under the density $f(\cdot)$ up to x
  - $\mathrm{E}[X] = \int_{-\infty}^{\infty} xf(x)dx.$ This weights x by the density.
  - Leading examples are normal, t distribution, F distribution.

Standard Normal Distribution Probabilities

- Compute $\Pr[a < X < b]$ as the **area under the probability density function** between a and b (this is an **integral**)
  - we cannot compute $\Pr[X = a]$ as it equals 0 for a continuous random variable.
- Standard normal example
  - $\Pr[X < 1.5] = \int_{-\infty}^{1.5} f(x)dx = \int_{-\infty}^{1.5} \frac{1}{\sqrt{2\pi}} \exp(-x^2/2)dx.$
  - $\Pr[X < 1.5] \simeq 0.9332$ by numerical methods
    - ★ as the integral has no closed form solution.
  - $\Pr[0.5 < X < 1.5] \simeq 0.2317$ similarly.



Appendix B.2: Probability Theory for the Sample Mean

- To get the properties of $\overline{X}$ we need results on sums of random variables.
- Statistical Independence
  - The random variables X and Y are **statistically independent** if the value taken by X is unaffected by the value taken by Y
  - Formally the joint probability of observing $X = x$ and $Y = y$ equals the product of the individual probabilities $\Pr[X = x] \times \Pr[Y = y]$.
- **Sums** of Random Variables: $X + Y$ and more generally $aX + bY$.
- Mean: $\mathrm{E}[aX + bY] = \mathrm{E}[aX] + \mathrm{E}[bY] = a\mathrm{E}[X] + b\mathrm{E}[Y]$.
- Variance: if X and Y are statistically independent then $\mathrm{Var}[aX + bY] = \mathrm{Var}[aX] + \mathrm{Var}[bY] = a^2\mathrm{Var}[X] + b^2\mathrm{Var}[Y]$.

Appendix B.2: Probability Theory for the Sample Mean

- **Sample mean** is the random variable $\bar{X} = \frac{1}{n}X_1 + \frac{1}{n}X_2 + \cdots + \frac{1}{n}X_n$.
- **Mean** of the sample mean
  - $\mathrm{E}[\bar{X}] = \frac{1}{n}\mu + \frac{1}{n}\mu + \cdots + \frac{1}{n}\mu = \mu$ if the X_i have common mean μ.
- **Variance** of the sample mean for statistically independent data

$$\begin{aligned}
\mathrm{Var}[\bar{X}] &= \mathrm{Var}\left[\tfrac{1}{n}X_1 + \tfrac{1}{n}X_2 + ... + \tfrac{1}{n}X_n\right] \\
&= \mathrm{Var}\left[\tfrac{1}{n}X_1\right] + \mathrm{Var}\left[\tfrac{1}{n}X_2\right] + \cdots + \mathrm{Var}\left[\tfrac{1}{n}X_n\right] \\
&= \left(\tfrac{1}{n}\right)^2 \mathrm{Var}\left[X_1\right] + \cdots + \left(\tfrac{1}{n}\right)^2 \mathrm{Var}\left[X_n\right] \\
&= \left(\tfrac{1}{n}\right)^2 \sigma^2 + \cdots + \left(\tfrac{1}{n}\right)^2 \sigma^2 \text{ if } X_i \text{ have common variance } \sigma^2 \\
&= \left(\tfrac{1}{n}\right) \sigma.
\end{aligned}$$

Behavior as sample size gets large

- Results can be obtained under weaker assumptions if the sample size is very large.
- Law of Large Numbers: $\bar{X} \to \mu$ as $n \to \infty$
  - this just requires just that X_i have common mean μ.
- Central Limit Theorem: $Z = (\bar{X} - \mu)/\sigma \sim N(0,1)$ as $n \to \infty$
  - this requires that X_i have common mean μ and common variance σ^2.
  - it can be generalized to $Z = (\bar{X} - \mathrm{E}[\bar{X}])/\mathrm{Var}[\bar{X}]$.

Appendix B.3: Probability Theory for two Related Variables

- Here we formally define conditional mean, conditional variance and correlation that are used in the main text.
- The **conditional probability** that event A happens given that B happens is the joint probability that both A and B happen divided by the probability that event B happens
  - $\Pr[A|B] = \frac{\Pr[A \text{ and } B]}{\Pr[B]}$
- Now translate to random variables X and Y.

Conditional probability

- $\Pr[X = x, Y = y]$ is the **joint probability** of Y and X.
- $\Pr[X = x]$ is the **marginal probability** of X
  - it is $\Pr[X = x] = \sum_y \Pr[X = x, Y = y]$ (sum over all y.values).
- Then the **conditional probability** of Y given X:
  $\Pr[Y = y|X = x] = \frac{\Pr[Y=y,X=x]}{\Pr[X=x]} = \frac{\text{Joint probability}}{\text{Marginal probability of } X}$.
- **Conditional mean** and **conditional variance** (of Y given X)
  - mean and variance with respect to conditional distribution
  - e.g. $E[Y|X = x] = \sum_y y \times \Pr[Y = y|X = x]$
  - the probability weighted average of y where the probability used is the conditional probability if Y given $X = x$.

Covariance and Correlation

- **Covariance** of X and Y
  - $\text{Cov}[X, Y] = \text{E}[(X - \mu_x)(Y - \mu_y)]$
  - calculate as $= \sum_x \sum_y (x - \mu_x)(y - \mu_y) \times \Pr[X = x, Y = y]$.
- **Correlation** of X and Y is the population analog of r_{xy}

$$\text{Cor}[X, Y] = \frac{\text{Cov}[X, Y]}{\sqrt{\text{Var}[X]}\sqrt{\text{Var}[Y]}}.$$

- X and Y are independent if $\Pr[X = x, Y = y] = \Pr[X = x] \times \Pr[Y = y]$.
- X and Y are uncorrelated if $\text{Cor}[X, Y] = 0$.
- Statistical independence implies uncorrelated.

Variance of sum of correlated variables

- We consider variance of a sum when random variables are correlated.

$$\begin{aligned}
\mathrm{Var}[Y_1+Y_2] &= \mathrm{E}[\{(Y_1+Y_2)-\mathrm{E}[(Y_1+Y_2)]\}^2] \\
&= \mathrm{E}[\{(Y_1+Y_2)-(\mu_1+\mu_2)\}^2] \\
&= \mathrm{E}[\{(Y_1-\mu_1)+(Y_2-\mu_2)\}^2] \\
&= \mathrm{E}[(Y_1-\mu_1)^2+(Y_2-\mu_2)^2+2(Y_1-\mu_1)(Y_2-\mu_2)] \\
&= \mathrm{E}[(Y_1-\mu_1)^2]+\mathrm{E}[(Y_2-\mu_2)^2] \\
&\qquad +2\mathrm{E}[(Y_1-\mu_1)(Y_2-\mu_2)] \\
&= \mathrm{Var}[Y_1]+\mathrm{Var}[Y_2]+2\mathrm{Cov}[Y_1,Y_2].
\end{aligned}$$

- More generally (using $\mathrm{Cov}[Y_i,Y_i]=\mathrm{Var}[Y_i]$)

$$\mathrm{Var}[\textstyle\sum_{i=1}^n Y_i]=\sum_i\sum_j \mathrm{Cov}[Y_i,Y_j].$$