ASIDE

Common univariate distributions

Distributions related to normal

1. Log-normal: If $X \sim N(\mu, \sigma^2)$ then $Y = e^X \sim \text{LogNormal}(\mu, \sigma^2)$

   - pdf given by
     $$\longrightarrow f_Y(y) = \frac{1}{y} f_X(\log y) = \frac{1}{y}\frac{1}{\sqrt{2\pi}\sigma} e^{-\frac{1}{2}\left(\frac{\log y - \mu}{\sigma}\right)^2}, \quad 0 < y < \infty$$
     $M_X(t) = \mathbb{E}[e^{tX}]$ ← normal $\Rightarrow M_X(t) = e^{\mu t + \frac{\sigma^2 t^2}{2}}$

   - Mean: $\text{E}Y = \text{E}e^X = M_X(t = 1) = e^{\mu + \frac{1}{2}\sigma^2}$ $\quad M_X(t=1) = e^{\mu + \frac{\sigma^2}{2}}$

     Note: $\text{E}Y = \text{E}e^X \geq e^{\text{E}X} = e^{\mu}$ by Jensen's inequality $\quad \mathbb{E}(\varphi(X)) \geq \varphi(\mathbb{E}X)$, $\varphi(x) = e^x$

   - shape is like that of a gamma distribution with $\alpha > 1$

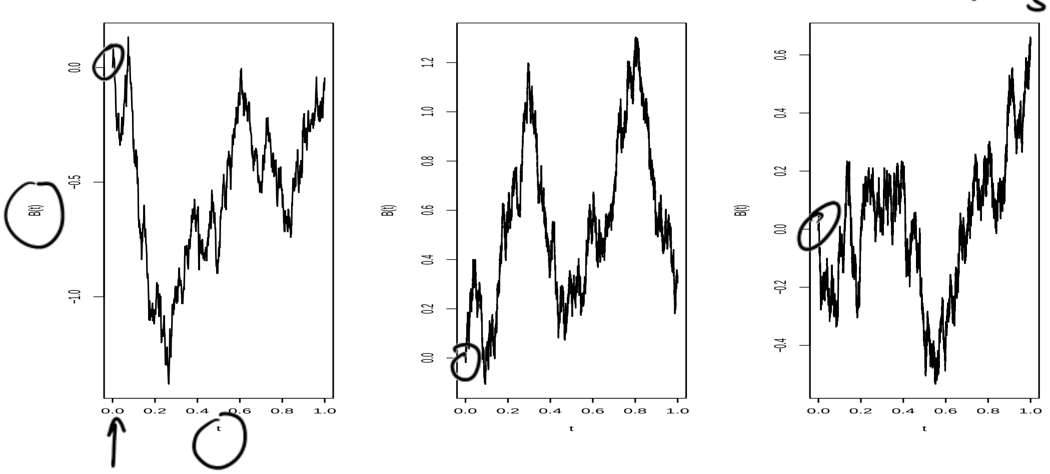
   - common in economics (e.g., assume log(income) is normal) and also as a failure time distribution

2. If $Z \sim N(0, 1)$ then $Z^2 \sim \chi^2_1$ (chi-squared 1 df)

3. Standard Browian motion $\{B(t) : t \geq 0\}$

   - a stochastic process where $B(t)$ represents an object's position at time t
   - $B(0) = 0$ and $B(t) \sim N(\mu = 0, \sigma^2 = t)$ for each $t > 0$
   - increments $B(s)$ and $B(t + s) - B(s)$ are independent, any $t > s \geq 0$
   - appears often in probability/statistics/economics

   s t t+s time



ASIDE

Common univariate distributions

Other continuous distributions

Each of the following has a location $-\infty < \mu < \infty$ and scale $\sigma > 0$ parameter

μ is the location Parameter
σ is the Scale Parameter

1. Cauchy

   - pdf given by

$$f_X(x) = \frac{1}{\pi\sigma\left(1 + \left(\frac{x-\mu}{\sigma}\right)^2\right)}, \quad -\infty < x < \infty$$

$g\left(\frac{x-\mu}{\sigma}\right)$

$\mathbb{E}X = \infty$
$\operatorname{Var} X = \infty$

   - a favorite counter-example distribution

   - no moments, no mgf

2. Logistic

   - pdf given by

$$f_X(x) = \frac{e^{-(x-\mu)/\sigma}}{\sigma\left(1 + e^{-(x-\mu)/\sigma}\right)^2}, \quad -\infty < x < \infty$$

   - underlies logistic regression

3. Double exponential

   - pdf given by

$$f_X(x) = \frac{1}{2\sigma} e^{-\frac{|x-\mu|}{\sigma}}, \quad -\infty < x < \infty$$

   - similar to normal but with heavier tails

Common univariate distributions

Other continuous distributions (cont'd)

4. Extreme value

   - pdf given by

$$f_X(x) = \frac{1}{\sigma} e^{(x-\mu)/\sigma} e^{-e^{(x-\mu)/\sigma}}, \quad -\infty < x < \infty$$

   - limiting distribution of record highs/lows

5. Pareto

   - pdf given by

$$f_X(x) = \frac{\alpha}{\beta \left(1 + \frac{x}{\beta}\right)^{\alpha+1}}, \quad 0 < x < \infty$$

   - $\alpha > 0$ is shape parameter, $\beta > 0$ is scale parameter

6. Several other distributions arise naturally in the statistical theory associated with normal distributions (more later), e.g.,

   - Student's t
   - Snedecor's F
   - χ_p^2

Families of distributions

Location-scale families of continuous random variables

- Already seen families with location and scale parameters (e.g., normal, Cauchy)

- *Definition:* Suppose $f_Z(z)$ is a given/fixed pdf. Then, the collection of pdfs of the form

$$f_X(x|\mu, \sigma) = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right),$$

new density function

for $-\infty < \mu < \infty$ and $\sigma > 0$, is the **location-scale family** with standard pdf f_Z.

- μ is location parameter; $\sigma > 0$ is scale parameter

- If $Z \sim f_Z(z)$ and $X = \mu + \sigma Z$ then $X \sim f_X(x|\mu, \sigma)$ as above

$$Z \longrightarrow X := \mu + \sigma Z \implies f_X(x) = f_Z\left(\frac{x-\mu}{\sigma}\right)$$

R.V. new

$$F_X(x) = \mathbb{P}(X \le x) = \mathbb{P}(\mu + \sigma Z \le x)$$

$$= \mathbb{P}\left(Z \le \frac{x-\mu}{\sigma}\right) = F_Z\left(\frac{x-\mu}{\sigma}\right)$$

- The standard pdf f_Z belongs to the family $f_X(x|\mu = 0, \sigma = 1)$ as above

$$f_X(x) = \frac{d}{dx} F_X(x) = \frac{d}{dx} F_Z\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right)$$

- Properties of the location-scale family can be obtained by studying the standard pdf f_Z

$$X = \mu + \sigma Z$$

e.g., $\mathrm{E}X = \mu + \sigma \mathrm{E}Z \qquad \mathrm{Var}(X) = \sigma^2 \mathrm{Var}(Z)$

- Also, the collection of pdfs of the form

$$f_X(x|\mu) = f_Z(x - \mu),$$

for $-\infty < \mu < \infty$, give the **location family** with standard pdf f_Z.

and pdfs of the form

$$f_X(x|\sigma) = \frac{1}{\sigma} f_Z\left(\frac{x}{\sigma}\right),$$

for $\sigma > 0$, give the **scale family** with standard pdf f_Z.

Families of distributions

Location-scale families (cont'd)

Example: standard normal $f_Z(z) = \frac{1}{\sqrt{2\pi}} e^{-z^2/2}$, $-\infty < z < \infty$, as the standard pdf

- The scale family ($\sigma > 0$) of normal distributions (with mean zero) is given by

$$X := \sigma Z \Rightarrow f_X(x) = \frac{1}{\sigma} f_Z\left(\frac{x}{\sigma}\right) = \frac{1}{\sigma}\frac{1}{\sqrt{2\pi}} e^{-\left(\frac{x}{\sigma}\right)^2/2}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-x^2}{2\sigma^2}}$$

$$X \sim N(0, \sigma^2)$$

- The location family ($-\infty < \mu < \infty$) of normal distributions (with variance 1) is given by

$$X := \mu + Z \Rightarrow f_X(x) = f_Z(x - \mu)$$

$$= \frac{1}{\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2}}$$

$$X \sim N(\mu, 1)$$

- The location-scale family ($-\infty < \mu < \infty$, $\sigma > 0$) of normal distributions is given by

$$X := \mu + \sigma Z$$

$$f_X(x) = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-\left(\frac{x-\mu}{\sigma}\right)^2}{2}}$$

$$= \frac{1}{\sqrt{2\pi}\,\sigma} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$$

Families of distributions

Location-scale families (cont'd)

$Z \sim Exp(1)$

Example: exponential distribution $f_Z(z) = e^{-z}$, $0 < z < \infty$, as the standard pdf

- The exponential scale family ($\sigma > 0$) is given by

$$X := \sigma Z \implies f_X(x) = \frac{1}{\sigma} f_Z\left(\frac{x}{\sigma}\right) = \frac{1}{\sigma} e^{-\frac{x}{\sigma}} \quad \forall x > 0$$

- The exponential location family ($-\infty < \mu < \infty$) is given by

$$X := \mu + Z \implies f_X(x) = f_Z(x - \mu) = e^{-(x-\mu)}, \quad \forall x > \mu$$

- The exponential location-scale family ($-\infty < \mu < \infty$, $\sigma > 0$) is given by

$$X := \mu + \sigma Z \implies f_X(x) = \frac{1}{\sigma} f_Z\left(\frac{x-\mu}{\sigma}\right) = \frac{1}{\sigma} e^{-\frac{(x-\mu)}{\sigma}}, \quad \forall x > \mu$$

STAT 542: Summary to date
Where we have been
& where we are headed

- Completed
  - Probability and random variables (definition, cdf, pdf/pmf)
  - Univariate distributions (definitions, expectation, transformations)
  - Families of univariate distributions

- Next
  - Multivariate distributions (Chapter 4, Sections 5.1, 5.2, 5.4)
    * extending the univariate ideas

      $\mathbb{P}(X=x \mid Y=y) \quad f_{X|Y}(x|y)$

    * conditional distributions and independence of r.v.s
    * hierarchical models
    * multivariate transformations
  - Order statistics

Multivariate distributions

Introduction

- Generally interested in more than one random variable at a time

   1. n observations of a single characteristic from some population

      $(X_1, \dots, X_n)$

   2. k different characteristics from a single individual

      One person Y_1 Y_2 Y_3

      age height gender

- *Notation:* $\underset{\sim}{X} = \boldsymbol{X} = (X_1, \dots, X_n)$ is an n-dimensional random vector (r.v.)

  In other words, $\boldsymbol{X}$ is a function from sample space $S \to \mathbb{R}^n$

  $X : \Omega \longrightarrow \mathbb{R}$ $(S \longrightarrow \mathbb{R})$ one-R.V.

  $\underset{\sim}{X} : \Omega \longrightarrow \mathbb{R}^n$, $\forall \omega \in \Omega$, $\underset{\sim}{X}(\omega) = \underbrace{(X_1(\omega), \dots, X_n(\omega))}_{\in \mathbb{R}^n}$

- General plan:

  use the bivariate case (X_1, X_2) $(n = 2)$ to derive results and then extend to any n

  For notational simplicity, in the bivariate case, we'll denote a pair of random variables as (X, Y) instead of (X_1, X_2)

Multivariate distributions

Another joint pdf example (i.e., continuous case)

Suppose (X, Y) have joint pdf $f(x, y) = 1/x$ for $0 < y < x < 1$

Multivariate distributions

Cumulative distribution functions

The joint cdf of X and Y is: $F_{X,Y}(x,y) = F(x,y) = P(X \leq x, Y \leq y)$, for $x, y \in \mathbb{R}$

1. rarely used in the discrete case but is computed as

$$F(x,y) = P(X \leq x, Y \leq y) = \sum_{x_1 \leq x, y_1 \leq y} f(x_1, y_1)$$

2. may be useful in the continuous case and is computed as

$$F(x,y) = P(X \leq x, Y \leq y) = \int_{-\infty}^{y} \int_{-\infty}^{x} f(s,t) ds dt$$

   where also in the continuous case

$$\frac{\partial^2 F(x,y)}{\partial x \partial y} = \frac{d \frac{dF(x,y)}{dy}}{dx} = f(x,y).$$

Properties: A function $F(x,y)$ is a cdf for some r.v. (X,Y) if and only if

1. $\lim\limits_{x \to -\infty} F(x,y) = \lim\limits_{y \to -\infty} F(x,y) = 0$

2. $\lim\limits_{x,y \to \infty} F(x,y) = 1$

3. Right continuous in each argument: $\lim\limits_{h \downarrow 0} F(x+h,y) = \lim\limits_{h \downarrow 0} F(x,y+h) = F(x,y)$

4. "nondecreasing" (the probability assigned to any rectangle is ≥ 0)

$$\begin{aligned} & P(x < X \leq x + \Delta_1, y < Y \leq y + \Delta_2) \\ = \ & F(x+\Delta_1, y+\Delta_2) - F(x+\Delta_1, y) - F(x, y+\Delta_2) + F(x,y) \geq 0 \end{aligned}$$

   for all $x, y \in \mathbb{R}$ and for any $\Delta_1, \Delta_2 > 0$

Multivariate distributions

Cumulative distribution functions (cont'd)

e.g., $F(x,y) = \begin{cases} 0 & \text{if } x+y < -1 \\ 1 & \text{if } x+y \geq -1 \end{cases}$ is *not* a valid cdf, though

Properties 1-3 hold, and $F(x,y)$ is nondecreasing in its arguments x or y (but "probabilities of rectangles" are not always non-negative).

Cdf example (continuous case): $f(x,y) = 4xy$, for $0 < x, y < 1$

$$F(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f(s,t)dsdt = \begin{cases} 1 & x \geq 1, y \geq 1 \\ y^2 & x \geq 1, 0 < y < 1 \\ x^2 & 0 < x < 1, y \geq 1 \\ x^2y^2 & 0 < x < 1, 0 < y < 1 \\ 0 & \text{otherwise} \end{cases}$$