Mathematics Bootcamp

Part I: Distributions

Jordan Bryan[1] Brian Cozzi[1] Michael Valancius[1] Graham Tierney[1] Becky Tang[1]

[1]Department of Statistical Science
Duke University

Graduate Orientation, August 2020

Outline

Random Variables

Random Variables

Definition: A *random variable* is a function from a sample space to an outcome space, e.g. the real numbers, integers, etc.

Suppose we want to perform a survey, run an experiment, do some quantitative study of a population of interest...

Let Ω be the set of all possible outcomes of a study.

Let $\omega \in \Omega$ be a particular outcome unit.

$Y = Y(\omega)$ is a function of ω and a random variable.

Random Variables - Example

The Experiment: 2 Dice are rolled together
The Sample Space: All pairs of numbers from 1 through 6
The Random Variable: The sum of the numbers

Cumulative Distribution Functions of Random Variables

Definition: The cumulative distribution function (CDF) or a random variable denoted by $F_X(x)$ is defined as:

$$F_X(x) = P_X(X \leq x); \quad \forall x$$

A function is a CDF if and only if the following are true:

- $\lim_{x \to -\infty} F(x) = 0$ and $lim_{x \to \infty} F(x) = 1$
- $F(x)$ is a non-decreasing function of x
- $F(x)$ is right continuous i.e. for every number x_0, $\lim_{x \to x_0^+} F(x) = F(x_0)$

An important implication of CDFs: *A random variable X is continuous if $F_X(x)$ is a continuous function of x. A random variable is discrete if $F_X(x)$ is a step function of x.*

Cumulative Distribution Functions of Random Variables - Example

If p denotes the probability of getting a head on any toss, and the experiment consists of tossing a coin until a head appears, then we define the random variable $X =$ the number of tosses required until a head. The CDF of this random variable is given as:

$$P(X \leq x) = \sum_{i=1}^{x} (1-p)^{i-1} p$$

Density and Mass Functions of Random Variables

Also related to the notion of a random variable is the concept of probability *density* and probability *mass* functions. Specifically, a *probability mass function* (PMF) for a discrete random variable is defined as:

$$f_X(x) = P(X = x); \forall x$$

and the *probability density function* (PDF) for a continuous random variable is defined as a function that satisfies the following relationship:

$$F_X(x) = \int_{-\infty}^{x} f_X(t)\mathrm{d}t; \forall x$$

Univariate Distributions

Discrete Distributions

A random variable X is *discrete* if the range of X, the sample space, is countable. In most situations, the random variable has integer valued outcomes

Some examples of discrete distributions:

- Binomial Distribution
- Poisson Distribution
- Negative Binomial Distribution
- Geometric Distribution

Binomial Distribution

This distribution counts the the number of successes in n independent trials all with the same fixed probability p of success

$$
\begin{aligned}
X &\sim \mathrm{Binomial}(n, p) \\
P(X = x) &= \frac{n!}{x!(n-x)!} p^x (1-p)^{n-x}, \quad x \in \{0, 1, \ldots, n\} \\
\mathbb{E}[X] &= np \\
\mathbb{V}[X] &= np(1-p)
\end{aligned}
$$

Poisson Distribution

This distribution is used for counting the number of events over some time horizon based on an intensity parameter λ

$$\begin{aligned} X &\sim \mathrm{Poisson}(\lambda) \\ P(X = x) &= \frac{e^{-\lambda}\lambda^x}{x!}, \quad x \in \{0, 1, 2, \ldots\} \\ \mathbb{E}[X] &= \mathbb{V}[X] = \lambda \end{aligned}$$

Negative Binomial Distribution

This distribution counts the the number successful trials k that occur before the rth failed trial, where each trial has fixed probability p of success

$$\begin{aligned} X &\sim \mathrm{NB}(r, p) \\ P(X = k) &= \binom{k + r - 1}{k} p^k (1 - p)^r, \quad k \in \{0, 1, 2, \ldots\} \\ \mathbb{E}[X] &= \frac{pr}{1 - p} \\ \mathbb{V}[X] &= \frac{pr}{(1 - p)^2} \end{aligned}$$

Is highly related to the Poisson and Gamma Distributions. And be careful about parameterizations!

Geometric Distribution

The Geometric distribution gives the probability that in a series of independent Bernoulli trials, each having probability of success p, we see the first success after k failures.

Here, the Geometric distribution is parameterized in terms of the number X Bernoulli failures the first success. You may also see the distribution parameterized in terms of number of trials.

$$
\begin{aligned}
X &\sim \text{Geom}(p) \\
P(X = k) &= (1-p)^k p, \quad k \in \{0, 1, 2, \ldots\} \\
\mathbb{E}[X] &= \frac{1-p}{p} \\
\mathbb{V}[X] &= \frac{-p}{p^2}
\end{aligned}
$$

Continuous Distributions

A random variable X is *continuous* if the range of X, the sample space, takes on an uncountably infinite number of values. In most instances the random variable has real-valued outcomes.

Some examples of Continuous Distributions

- Normal Distribution
- Chi-Squared Distribution
- Exponential Distribution
- Gamma Distribution
- Inverse-Gamma Distribution
- Student-t Distribution
- F Distribution
- Beta Distribution

Normal Distribution

A random variable $X \sim \mathrm{Normal}(\mu, \sigma^2)$ with PDF:

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

We also sometimes express this in terms of a *precision* parameter, rather than a variance, $X \sim \mathrm{Normal}(\mu, \phi^{-1})$ which becomes useful when performing Bayesian inference. If $Z \sim \mathrm{Normal}(0, 1)$ then the distribution of Z is standard normal.

Chi-Squared Distribution

If Z_1, $Z_2, \ldots, Z_k$ are independent, standard normal random variables, then

$$\sum_{j=1}^{k} Z_j^2 \sim \chi_k^2$$

follows a Chi-Squared distribution with k degrees of freedom. This is a special case of the Gamma distribution, discussed on the next slide.

Gamma Distribution

A random variable $X \sim \mathrm{Gamma}(\alpha, \beta)$ with PDF:

$$f_X(x) = \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} \exp\{-x\beta\}$$

$$\mathbb{E}[X] = \frac{\alpha}{\beta}$$

$$\mathbb{V}[X] = \frac{\alpha}{\beta^2}$$

$$\alpha, \beta > 0$$

$$x \in (0, \infty)$$

Note: this is the shape-rate parameterization. You may also see the shape-scale parameterization, with scale $\theta = 1/\beta$.

Gamma Distribution - Important Properties

Here are some important tricks that will be useful in **711** and **601**

- if $\alpha = 1$ and then $X \sim \text{Exponential}(\lambda = \beta)$
- if $\alpha = \frac{\nu}{2}$ and $\beta = \frac{1}{2}$ then $X \sim \chi^2_\nu$
- if $X \sim \text{Gamma}(\alpha_1, \beta)$ and $Y \sim \text{Gamma}(\alpha_2, \beta)$ then $X + Y \sim \text{Gamma}(\alpha_1 + \alpha_2, \beta)$
- if $X \sim \text{Gamma}(k, \theta)$, then $\frac{1}{X} \sim \text{Inverse} - \text{Gamma}(k, \frac{1}{\theta})$

Student's-t Distribution

A random variable T follows a Student's-t distribution if

$$T = \frac{Z}{\sqrt{V/\nu}},$$
$$Z \sim N(0, 1),$$
$$V \sim \chi^2_\nu$$

and Z and V are independent.

F Distribution

A random variable X follows a F-distribution with numerator degrees of freedom ν_1 and denominator degrees of freedom ν_2 if

$$X = \frac{V_1/\nu_1}{V_2/\nu_2}$$

where V_1 and V_2 are independent chi-squared random variables with degrees of freedom equal to ν_1 and ν_2 respectively.

Beta Distribution

A random variable $X \sim \mathrm{Beta}(\alpha, \beta)$ with PDF:

$$
\begin{aligned}
f_X(x) &= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \\
\mathbb{E}[X] &= \frac{\alpha}{\alpha+\beta} \\
\mathbb{V}[X] &= \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)} \\
&\alpha, \beta > 0 \\
&x \in (0,1)
\end{aligned}
$$

Very useful for eliciting probability distributions for proportions.

Cool distributional relationships

Exponential Families

Exponential Families

Random variables belong to the exponential family if their PMFs/PDFs can be expressed in the form:

$$f_X(x|\theta) = h(x)c(\theta)\exp\left\{\sum_{i=1}^{k}\omega_i(\theta)t_i(x)\right\}$$

Where:

$$h(x) \geq 0$$
$$c(\theta) \geq 0$$
$$\omega_1(\theta), \ldots, \omega_k(\theta) \in \mathbb{R}$$
$$t_1(x), \ldots t_k(x) \in \mathbb{R}$$

We will show later some of the convenient properties of the exponential family distributions.

Exponential Families - Example

Consider the binomial PMF for a random variable $X \sim \text{Binomial}(n, p)$

$$P(X = x) = \binom{n}{x} p^x (1-p)^{n-x}$$

This is an exponential family PMF. We can show this by re-expressing terms:

$$P(X = x) = \binom{n}{x} (1-p)^n \exp\left\{ x \log\left(\frac{p}{1-p}\right) \right\}$$

$$h(x) = \binom{n}{x} \mathbb{I}_{x=0,\ldots,n}$$

$$c(p) = (1-p)^n$$

$$\omega_1(p) = \log\left(\frac{p}{1-p}\right)$$

$$t_1(x) = x$$

Exponential Families - Exercise

Consider the following normal PDF for $X \sim \mathrm{Normal}(\mu, \sigma^2)$

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

Show that this is an exponential family PDF

Exponential Families - Exercise Cont.

Consider the following PDF

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

Expanding the exponential yields the following

$$
\begin{aligned}
f_X(x|\mu,\sigma^2) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{-\frac{\mu^2}{2\sigma^2}\} \exp\{-\frac{x^2}{2\sigma^2} + \frac{\mu x}{\sigma^2}\} \\
h(x) &= 1 \\
c(\mu,\sigma) &= \frac{1}{\sqrt{2\pi\sigma^2}} \exp\{\frac{-\mu^2}{\sigma^2}\}; \mu \in \mathbb{R} \quad \sigma^2 > 0 \\
\omega_1(\mu,\sigma) &= \frac{1}{\sigma^2} \quad \omega_2 = \frac{\mu}{\sigma^2} \\
t_1(x) &= \frac{-x^2}{2} \quad t_2(x) = x
\end{aligned}
$$

Transformations of Random Variables

Transformations of Random Variables using the Change of Variables Formula

Assume that X has a pdf $f_X(x)$ and that $Y = g(X)$ where g is a monotone function. Suppose that $f_X(x)$ is continuous on $\mathcal{X}$, and that g^{-1} has a continuous derivative on $\mathcal{Y}$ where $\mathcal{X}, \mathcal{Y}$ are such that $\mathcal{X} = \{x : f_X(x) > 0\}$ and $\mathcal{Y} = \{y : y = g(x)\}$. Then the pdf of Y is given as follows:

$$f_Y(y) = f_X(g^{-1}(y))|\frac{\mathrm{d}}{\mathrm{d}y}g^{-1}(y)|$$

Transformations of Random Variables - Example

Assume that $X \sim f_X(x) = 1$ i.e. $X \sim \mathrm{Uniform}(0,1)$. Furthermore, $Y = -\log(X)$. What is the PDF of Y?

First note that $g(X) = Y = -\log(X) \to g^{-1}(Y) = e^{-Y}$. Therefore, using the formulation from earlier:

$$f_Y(y) = 1 \cdot |-e^{-y}| = e^{-y}$$
$$Y \sim \mathrm{Exponential}(\lambda = 1)$$

Transformations of Random Variables - Exercise

Assume that $X \sim \mathrm{Normal}(0, 1)$. Let $Y = X^2$. What is the distribution of Y?

The PDF of the standard normal distribution is given as follows:

$$f_X(x) = \frac{1}{\sqrt{2\pi}} \exp\{-\frac{x^2}{2}\}$$

Transformations of Random Variables - Exercise Cont.

Consider that $Y = g(X) = X^2 \to g^{-1}(Y) = \pm\sqrt{Y}$. Hence, consider that we can partition the support of X into two pieces $S_1 = (-\infty, 0)$ and $S_2 = (0, \infty)$ where the function $g(X)$ is monotone. Note that $\mathcal{Y} = (0, \infty)$. Use the change of variables formulation over the two partitions and sum:

$$
\begin{aligned}
f_Y(y) &= \frac{1}{\sqrt{2\pi}} \exp\{\frac{-(-\sqrt{Y})^2}{2}\}|-\frac{1}{2\sqrt{y}}| + \frac{1}{\sqrt{2\pi}} \exp\{\frac{-(\sqrt{Y})^2}{2}\}|\frac{1}{2\sqrt{y}}| \\
&= \frac{1}{\sqrt{2\pi}} \frac{1}{\sqrt{Y}} \exp\{\frac{-y}{2}\}
\end{aligned}
$$

Hence, we get that $Y \sim \chi^2_{df=1}$

Multivariate Distributions

Random Vectors

If we have d random variables $X_1, X_2, \ldots, X_d$, each defined on the real line, we can write them as the d dimensional column vector

$$\mathbf{X} = (X_1, \cdots X_d)^T$$

which we call a d-dimensional **random vector**. The joint distribution function of the random vector $\mathbf{X}$ is

$$\begin{aligned} F_X(\mathbf{x}) &= F_X(x_1, \ldots, x_d) \\ &= P(X_1 \leq x_1, \ldots, X_d \leq x_d) \\ &= P(\mathbf{X} \leq \mathbf{x}) \end{aligned}$$

If F_X is absolutely continuous, then the joint density function f_X of $\mathbf{X}$ is

$$f_X(\mathbf{x}) = f_X(x_1, \ldots, x_d) = \frac{\partial^d F_X(x_1, \ldots, x_d)}{\partial x_1 \cdots \partial x_d}$$

Random Vectors

To find the marginal density of a subset of the d variables, you can just integrate the others out. For example, if we have a joint bivariate density $f_{X_1,X_2}(x_1, x_2)$, then

$$f_{X_1}(x_1) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2) dx_2 \qquad f_{X_2}(x_2) = \int_{-\infty}^{\infty} f_{X_1,X_2}(x_1, x_2) dx_1$$

The components of a random vector $\mathbf{X}$ are **independent** if the joint distribution function is a product of the marginal distribution functions

$$F_X(\mathbf{x}) = \prod_{i=1}^{d} F_i(x_i)$$

In addition, the joint density is the product of marginals

$$f_X(\mathbf{x}) = \prod_{i=1}^{d} f_i(x_i)$$

Expectation and Covariance

If $\mathbf{X}$ is a random vector with values in $\mathbb{R}^d$, then its expected value is given by the d dimensional vector

$$\mu_X = E(\mathbf{X}) = (E(X_1), \cdots, E(X_d)) = (\mu_1, \cdots, \mu_d)^T$$

and the $d \times d$ **covariance matrix** of $\mathbf{X}$ is

$$\begin{aligned}
\boldsymbol{\Sigma}_{XX} &= \text{cov}(\mathbf{X}, \mathbf{X}) \\
&= E[(\mathbf{X} - \mu_X)(\mathbf{X} - \mu_X)^T] \\
&= E[(X_1 - \mu_1, \cdots, X_d - \mu_d)(X_1 - \mu_1, \cdots, X_d - \mu_d)^T] \\
&= \begin{pmatrix} \sigma_1^2 & \sigma_{12} & \dots & \sigma_{1d} \\ \sigma_{21} & \sigma_2^2 & \dots & \sigma_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \sigma_{d1} & \sigma_{d2} & \dots & \sigma_d^2 \end{pmatrix}
\end{aligned}$$

Correlation Matrix

The **correlation matrix** of $\mathbf{X}$ can be obtained by from $\boldsymbol{\Sigma}_{XX}$ by dividing the ith row by σ_i and the jth column by σ_j. The $d \times d$ matrix is then

$$P_{XX} = \begin{pmatrix} 1 & \rho_{12} & \dots & \rho_{1d} \\ \rho_{21} & 1 & \dots & \rho_{2d} \\ \vdots & \vdots & \ddots & \vdots \\ \rho_{d1} & \rho_{d2} & \dots & 1 \end{pmatrix}$$

where

$$\rho_{ij} = \rho_{ji} = \begin{cases} \frac{\sigma_{ij}}{\sigma_i \sigma_j} & i \neq j \\ 1 & \text{otherwise} \end{cases}$$

is the pairwise correlation coefficient between X_i and X_j. The correlation coefficient will always lie between -1 and 1 and is a measure of association between X_i and X_j.

Linear Functions of Random Vectors

If $\mathbf{Y}$ is a linear function of $\mathbf{X}$ such that

$$\mathbf{Y} = \mathbf{A}\mathbf{X} + \mathbf{b}$$

the mean vector and covariance matrix of $\mathbf{Y}$ is given by

$$\mu_Y = \mathbf{A}\mu_x + \mathbf{b}$$
$$\mathbf{\Sigma}_{YY} = \mathbf{A}\mathbf{\Sigma}_{\mathbf{XX}}\mathbf{A}^T$$

Multivariate Normal Distribution

The form of the multivariate normal looks similar to that of the univariate normal. A random d vector $\mathbf{X}$ follows a multivariate normal distribution with mean vector μ and positive definite symmetric covariance matrix $\mathbf{\Sigma}$ if it has the density function

$$f(\mathbf{x}|\mu, \mathbf{\Sigma}) = (2\pi)^{-d/2}|\mathbf{\Sigma}|^{-1/2}e^{-\frac{1}{2}(\mathbf{x}-\mu)^T\mathbf{\Sigma}^{-1}(\mathbf{x}-\mu)}$$

Notation: we denote a d dimensional normal distribution as

$$\mathbf{X} \sim N_d(\mu, \mathbf{\Sigma})$$

Multivariate Normal Distribution

The **Mahalanobis distance** from $\mathbf{x}$ to μ is given by the quadratic form

$$\Delta = \sqrt{(\mathbf{x} - \mu)^T \mathbf{\Sigma}^{-1} (\mathbf{x} - \mu)}$$

An important result is that a random vector $\mathbf{X}$ follows a multivariate distribution if and only if every linear function of $\mathbf{X}$ follows a univariate normal distribution.

In linear models, we often assume that $\mathbf{\Sigma} = \sigma^2 \mathbf{I_d}$, in which case the density function reduces to

$$f(\mathbf{x}|\mu, \sigma) = (2\pi\sigma)^{-d/2} e^{-\frac{1}{2}(\mathbf{x}-\mu)^T(\mathbf{x}-\mu)}$$