Mathematics Bootcamp

Part III: Probability

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

Moments

Expectations and Variances of Random Variables

The expectation of any random variable can be computed as follows:

- $\mathbb{E}[g(X)] = \int_{-\infty}^{\infty} g(x) f_X(x) \mathrm{d}x$ when X is continuous
- $\mathbb{E}[g(X)] = \sum_{x \in \mathcal{X}} g(x) f_X(x) = \sum_{x \in \mathcal{X}} g(x) \mathbb{P}(X = x)$ when X is discrete

We are often interested in the expectation of a random variable itself, so $g(x) = x$.
The variance can be computed using the expectations as follows:

$$\mathbb{V}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$$

You will often need to do some calculus to find each of these quantities

Poisson Distribution - Exercise

Prove that $\mathbb{E}[X] = \lambda$ if $X \sim \mathrm{Poisson}(\lambda)$

Poisson Distribution - Exercise Cont.

We need to compute the following:

$$\mathbb{E}[X] = \sum_{x=0}^{\infty} x \frac{e^{-\lambda}\lambda^{x}}{x!}$$
$$= \lambda e^{-\lambda} \sum_{x=1}^{\infty} \frac{\lambda^{x-1}}{(x-1)!}$$

Now recognize the following result from the Taylor series expansion on $\exp\{y\} = \sum_{i=0}^{\infty} \frac{y^i}{i!}$. Use this result with a clever substitution:

$$\lambda \exp\{-\lambda\} \exp\{\lambda\} = \lambda$$

Properties of Expectations and Variances

Let X be a random variable and let a be a scalar constant, then:

$$\mathbb{E}[aX] = a\mathbb{E}[X]$$
$$\mathbb{V}[aX] = a^2\mathbb{V}[X]$$

Expectation is a linear operator, so if X and Y are two random variables, then $\mathbb{E}[X + Y] = \mathbb{E}[X] + \mathbb{E}[Y]$.

For variance,

$$\mathbb{V}[X \mp Y] = \mathbb{V}[X] + \mathbb{V}[Y] \mp 2\mathbb{C}(X, Y),$$

where $\mathbb{C}(X, Y)$ denotes the covariance between X and Y. These extend to multivariate random variables as well.

Total Expectation and Total Variance Laws

In many examples, you are interested in marginal moments from conditional distributions (more on these distributions coming up). Your first option of course is to find the joint distribution, do some marginalization and then integrate, but I do not like calculus so **instead**:

$$\mathbb{E}[Y] = \mathbb{E}[\mathbb{E}[Y|X]]$$
$$\mathbb{V}[Y] = \mathbb{V}[\mathbb{E}[Y|X]] + \mathbb{E}[\mathbb{V}[Y|X]]$$

Total Expectation and Total Variance Laws - Example

Assume that we have the following relationship:

$$X|N \sim \text{Binomial}(N, p)$$
$$N \sim \text{NegativeBinomial}(\tau = \frac{1}{1+\beta}, r = 1)$$

Find $\mathbb{E}[X]$ and $\mathbb{V}[X]$

Tip: $\mathbb{E}[N] = \frac{r\tau}{1-\tau}$ and $\mathbb{V}[N] = \frac{\tau r}{(1-\tau)^2}$

Total Expectation and Total Variance Laws - Example Cont.

First, we iterate to find the expectation

$$\begin{aligned}\mathbb{E}[X] &= \mathbb{E}[\mathbb{E}[X|N]] \\ &= \mathbb{E}[Np] \\ &= p\frac{\frac{1}{1+\beta}}{1-\frac{1}{1+\beta}} \\ &= \frac{p}{\beta}\end{aligned}$$

Next, we proceed with finding the variance

$$\begin{aligned}\mathbb{V}[X] &= \mathbb{E}[\mathbb{V}[X|N]] + \mathbb{V}[\mathbb{E}[X|N]] \\ &= \mathbb{E}[Np(1-p)] + \mathbb{V}[Np] \\ &= \frac{p(1-p)}{\beta} + p^2\frac{1+\beta}{\beta^2}\end{aligned}$$

Total Expectation and Total Variance Laws - Exercise

$$X|P \sim \mathrm{Binomial}(n, P)$$
$$P \sim \mathrm{Beta}(a, b)$$

Find the $\mathbb{E}[X]$ and $\mathbb{V}[X]$
Tip:

$$\mathbb{E}[P] = \frac{a}{a+b}$$
$$\mathbb{V}[P] = \frac{ab}{(a+b)^2(a+b+1)}$$

Total Expectation and Total Variance Laws - Exercise Cont.

We can start by finding the marginal expectation first:

$$\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X|P]] = \mathbb{E}[nP] = n\mathbb{E}[P] = n\frac{a}{a+b}$$

And then the marginal variance:

$$\begin{aligned}
\mathbb{V}[X] &= \mathbb{V}[\mathbb{E}[X|P]] + \mathbb{E}[\mathbb{V}[X|P]] \\
&= \mathbb{V}[nP] + \mathbb{E}[nP(1-P)] \\
&= n^2\mathbb{V}[P] + n\mathbb{E}[P - P^2] \\
&= n^2\frac{ab}{(a+b)^2(a+b+1)} + n\frac{a}{a+b} \\
&- n(\frac{ab}{(a+b)^2(a+b+1)}) - n(\frac{a}{a+b})^2 \\
&= n\frac{ab(a+b+n)}{(a+b)^2(a+b+1)}
\end{aligned}$$

Moment Generating Functions

- The moment generating function (MGF)

$$M_x(t) = \mathbb{E}[e^{tX}]$$

  uniquely defines the distribution of a random variable
- So for X discrete, $M_x(t) = \sum_{x \in S_x} e^{tx} P(X = x)$
- And for X continuous, $M_x(t) = \int_{x \in S_x} e^{tx} f_X(x) dx$
- Why do we care? Besides the uniqueness property, MGFs are extremely helpful for determining distributions of sums of independent random variables

MGF Example: Gamma Distribution

Suppose $X \sim Gamma(\alpha, \beta)$. Then

$$
\begin{aligned}
M_x(t) &= \int_0^\infty e^{xt} \frac{\beta^\alpha}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x^{\alpha-1} e^{-x(\beta-t)} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \int_0^\infty x \cdot x^{(\alpha-1)-1} e^{-x(\beta-t)} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha-1)}{(\beta-t)^{\alpha-1}} \times \\
&\quad \int_0^\infty x \cdot \frac{(\beta-t)^{\alpha-1}}{\Gamma(\alpha-1)} x^{(\alpha-1)-1} e^{-x(\beta-t)} dx \\
&= \frac{\beta^\alpha}{\Gamma(\alpha)} \cdot \frac{\Gamma(\alpha-1)}{(\beta-t)^{\alpha-1}} \mathbb{E}[Y], \quad Y \sim Ga(\alpha-1, \beta-t)
\end{aligned}
$$

MGF Example: Gamma Distribution

Recalling the expectation of a Gamma and simplifying terms:

$$\begin{aligned} M_x(t) &= \frac{\beta^\alpha}{(\alpha-1)!} \frac{(\alpha-2)!}{(\beta-t)^{\alpha-1}} \cdot \frac{\alpha-1}{\beta-t} \\ &= \left(\frac{\beta}{\beta-t}\right)^\alpha, \quad \text{for } t < \beta \end{aligned}$$

MGF of Chi-Squared

Rather than brute-force deriving the MGF for the χ^2, we can use the MGF of a Gamma random variable. We previously noted that if $Y \sim Gamma(k/2, 1/2)$, then Y is also distributed as χ^2_k. So the MGF of $Y \sim \chi^2_k$ is

$$M_y(t) = \left(\frac{1/2}{1/2 - t}\right)^{k/2} = \left(\frac{1}{1 - 2t}\right)^{k/2} = (1 - 2t)^{-k/2}$$

Probability Theory

Distribution Functions for Multivariate Random Variables

There are three types of distribution functions that we will cover:

- Joint Distribution
- Marginal Distribution
- Conditional Distribution

Joint Distribution - Bivariate Case

Joint PDF: A function $f(x, y)$ from $\mathbb{R}^2 \to \mathbb{R}$ is called a joint PDF of the random vector (X, Y) if for every $A \subset \mathbb{R}^2$

$$\mathbb{P}((X, Y) \in A) = \int_A \int f_{X,Y}(x, y) \mathrm{d}x\mathrm{d}y$$

Joint PMF: The function $f(x, y)$ from $\mathbb{R}^2 \to \mathbb{R}$ defined by $f_{X,Y}(x, y) = \mathbb{P}(X = x, Y = y)$ is the joint PMF of X, Y. Then for every $A \subset \mathbb{R}^2$

$$\mathbb{P}((X, Y) \in A) = \sum_{(x,y) \in A} f_{X,Y}(x, y)$$

Joint Distribution - Exercise

Assume that X and Y have the joint PDF:

$$f_{X,Y}(x,y) = 4xy$$
$$0 < x < 1$$
$$0 < y < 1$$

Find $P(Y < X)$ * Show that this is proper *

Joint Distribution - Exercise Cont.

We can set up the double integral required for this probability as follows:

$$
\begin{aligned}
p(Y < X) &= \int_0^1 \int_0^x 4xy \mathrm{d}y \mathrm{d}x \\
&= \int_0^1 [4x\frac{y^2}{2}]\Big|_0^x \mathrm{d}x \\
&= \int_0^1 2x^3 \mathrm{d}x = \frac{1}{2}
\end{aligned}
$$

Marginal Distribution

Given the joint PDF or joint PMF, you can find the marginal PDF or PMF:

Marginal PDF:

$$f_X(x) = \int_Y f_{X,Y}(x, y)\mathrm{d}y$$

Marginal PMF:

$$f_Y(y) = \sum_x f_{X,Y}(x, y)$$

Conditional Probability and Independence

Starting with something familiar. Consider two events A and B with the sample space Ω.

$$\mathbb{P}(A|B) = \frac{\mathbb{P}(A \cap B)}{\mathbb{P}(B)}$$

Furthermore, consider the following notion of independence for the same two events. A and B are independent if:

$$\mathbb{P}(A \cap B) = \mathbb{P}(A)\mathbb{P}(B)$$

Conditional Probability and Independence - Continued

Conditional Probability for more than two events. Let $A_1, A_2, \ldots$ be a partition of the sample space and let B be any set, then for $i = 1, 2, \ldots$:

$$\mathbb{P}(A_i|B) = \frac{\mathbb{P}(B|A_i)\mathbb{P}(A_i)}{\sum_{i=1}^{\infty} \mathbb{P}(B|A_j)\mathbb{P}(A_j)}$$

We can similarly extend the definition of independence to cases with more then two events. A collection of events $A_1, \ldots, A_n$ are considered mutually independent if for any subcollection $A_{i_1}, \ldots, A_{i_K}$ we have that:

$$\mathbb{P}(\cap_{j=1}^{K} A_{i_j}) = \prod_{j=1}^{K} \mathbb{P}(A_{i_j})$$

Conditional Distribution

Assume that $X, Y \sim f_{X,Y}(x, y)$, then we can employ Bayes' rule for distributions:

$$f_{X|Y}(x|y) = \frac{f_{X,Y}(x, y)}{f_Y(y)}$$

Conditional Probability - Example

In morse code, information is represented as dots and dashes. Assume the following:

$$\mathbb{P}(dot \quad sent) = \frac{3}{7}; \quad \mathbb{P}(dash \quad sent) = \frac{4}{7}$$

Furthermore, we also know that $\mathbb{P}(dot \quad received | dot \quad sent) = \frac{7}{8}$. Find $\mathbb{P}(dot \quad sent | dot \quad received)$.

Conditional Probability - Example Cont.

In order to use Bayes Rule, we first need $\mathbb{P}(dot \quad received)$.

$$\mathbb{P}(dot \quad received) = \mathbb{P}(dot \quad received \cap dot \quad sent) +$$

$$\mathbb{P}(dot \quad received \cap \quad dash \quad sent) \qquad = \frac{7}{8}\frac{3}{7}$$

$$+ \left(\frac{1}{8}\right)\left(\frac{4}{7}\right) = \frac{25}{26}$$

Applying Bayes Rule:

$$\mathbb{P}(dot \quad sent | dot \quad received) = \frac{\mathbb{P}(dot \quad sent \cap dot \quad received)}{\mathbb{P}(dot \quad sent)}$$

$$= \frac{\left(\frac{7}{8}\right)\left(\frac{3}{7}\right)}{\frac{25}{26}}$$

Conditional Probability - Exercise

In the population the probability of an infectious disease is $\mathbb{P}(D) = 0.01$. The probability of testing positive if the disease is present is $\mathbb{P}(+|D) = 0.95$. The probability of a negative test given the disease is not present is $\mathbb{P}(-|ND) = 0.95$. What is the probability of the disease being present if the test is positive i.e. $\mathbb{P}(D|+)$?

Conditional Probability - Exercise Cont.

First find the probability of a positive test:

$$\begin{aligned}\mathbb{P}(+) = \mathbb{P}(+|D)P(D) + \mathbb{P}(+|ND)P(ND) &= 0.01 \cdot 0.95 + 0.05 \cdot 0.99 \\ &= 0.059\end{aligned}$$

Next, we can invoke Bayes Rule:

$$\mathbb{P}(D|+) = \frac{\mathbb{P}(D \cap +)}{\mathbb{P}(+)} = \frac{0.01 \cdot 0.95}{0.059} \approx 0.161$$

Conditional Distributions - Exercise

Assume that (X, Y) are a continuous random vector with joint pdf given by:

$$f_{X,Y}(x, y) = \exp\{-y\} \;\; 0 < x < y < \infty$$

Find the marginal distribution of X and the conditional distribution $Y|X$

Conditional Distributions - Example Cont.

We start by finding the marginal distribution of X:

$$f_X(x) = \int_x^\infty \exp\{-y\}\mathrm{d}y = e^{-x}$$
$$X \sim \mathrm{Exponential}(\lambda = 1)$$

Now use the results on conditional distributions given earlier, to find:

$$f_{Y|X}(y|x) = \frac{f_{X,Y}(x,y)}{f_X(x)} = \frac{\exp\{-y\}}{\exp\{-x\}}\mathbb{I}(x < y)$$

Independence - Example

Consider an experiment of tossing two dice. The sample space is therefore:

$$\Omega = \{(1,1),(1,2),\dots(1,6),(2,1),\dots,(2,6),\dots,(6,6)\}$$

Further, we define the events:

$$A = \{\text{doubles appear}\}$$
$$B = \{\text{the sum is between 7 and 10}\}$$
$$C = \{\text{the sum is 2 or 7 or 10}\}$$

Are the events A, B, C mutually independent?

Independence - Example Cont.

Note that the following can be found by enumeration:

$$\mathbb{P}(A) = \frac{1}{6}; \;\; \mathbb{P}(B) = \frac{1}{2}; \;\; \mathbb{P}(C) = \frac{1}{3}$$

Furthermore:

$$\begin{aligned} \mathbb{P}(A \cap B \cap C) &= \mathbb{P}(\text{sum is 10, comprised of doubles}) = \frac{1}{36} \\ &= \mathbb{P}(A)\mathbb{P}(B)\mathbb{P}(C) = \frac{1}{6} \cdot \frac{1}{2} \cdot \frac{1}{3} \end{aligned}$$

But notice that $\mathbb{P}(B \cap C) = \frac{11}{36} \neq \mathbb{P}(B)\mathbb{P}(C)$. Therefore we do not have pairwise independence and hence claims of mutual independence cannot be made.

Independence - Exercise

Consider the following sample sample that consists of the 3! permutations of $\{a, b, c\}$ along with triples of each letter:

$$\Omega = \{aaa, bbb, ccc, abc, bca, cba, acb, bac, cab\}$$

Each element in Ω is assumed to have probability $\frac{1}{9}$. Define the event A_i:

$$A_i = \{i^{th} \;\; place \;\; in \;\; the \;\; triple \;\; is \;\; occupied \;\; by \;\; a\};$$
$$i = 1, 2, 3$$
$$\mathbb{P}(A_i) = \frac{1}{3}$$

Are the events A_i mutually independent?

Independence - Exercise Cont.

Pairwise independence is satisfied:

$$\mathbb{P}(A_1 \cap A_2) = \mathbb{P}(A_1 \cap A_3) = \mathbb{P}(A_2 \cap A_3) = \frac{1}{9}$$

But the joint event:

$$\mathbb{P}(A_1 \cap A_2 \cap A_3) = \frac{1}{9} \neq \mathbb{P}(A_1)P(A_2)\mathbb{P}(A_3)$$

Hence, the events are **not** mutually independent

Conditional Independence

True independence is pretty rare in most applications, so we generally rely on *conditional independence*.

Suppose we have three random variables Y_1, Y_2, Y_3 that we believe are *"independent and identically distributed (iid)"*. Does our knowledge about the value of one inform our knowledge about another?

$$Pr(Y_1 = y_1 \mid Y_2 = y_2, Y_3 = y_3) = P(Y_1 = y_1)?$$

Conditional Independence

Suppose Y_1, Y_2, Y_3 are *Conditionally Independent* given some parameter vector θ. This means that

$$Pr(Y_1 = y_1 \mid \theta, Y_2 = y_2, Y_3 = y_3) = Pr(Y_1 = y_1 \mid \theta)$$

Now we can say,

$$P(Y_1, Y_2, Y_3 \mid \theta) = P(Y_1 \mid \theta)P(Y_2 \mid \theta)P(Y_3 \mid \theta)$$

Suppose we want to find the value of θ that makes these data most likely... See MLE