STAT 110 Refresher

Ricky Truong (rickytruong@college.harvard.edu),
Jesse Brodtman (jbrodtman@college.harvard.edu)

1 Introduction

1.1 Logistics

Welcome! The goal of this session is to review the important concepts from STAT 110 ("Introduction to Probability"), especially as they apply to higher-level courses like STAT 111 ("Introduction to Statistical Inference").

Specifically, we aim for this session to be interactive, well-paced, inclusive, and fun! Please do not hesitate to reach out if you have any questions/feedback!

1.2 Notation

Unless noted otherwise, we use the following conventions:

- Capital letters for random variables and events. E.g., $X \sim \text{Bern}(p)$, $P(A) = 0.5$.
- Lower-case letters for crystallizations (i.e., realized values). E.g., $P(X = x) = p^x(1-x)^{1-x}$.
- Double right arrows for implications. E.g., $x + 2 = 5 \implies x = 3$.
- $\log(x)$ for the natural logarithm of x. E.g., $\log(e) = 1$.
- $\mathbb{1}\{A\}$ for indicators. E.g., $\mathbb{E}[\mathbb{1}\{A\}] = P(A)$.

1.3 Join GUSH!

The Group for Undergraduates in Statistics at Harvard (GUSH) hosts events throughout the school year, open to all students! Join our mailing list for other workshops as well as social and professional opportunities: http://www.gushclub.org.

2 Mathematics

- $e^x = \sum_{n=0}^{\infty} \frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \dots$ by Taylor series.
- $e^x = \lim_{n\to\infty} \left(1 + \frac{x}{n}\right)^n$ by compound interest.
- $\sum_{n=0}^{\infty} x^n = 1 + x + x^2 + \dots = \frac{1}{1-x}$ for $|x| < 1$ by infinite geometric series.
- $\sum_{k=0}^{n-1} ar^k = a + ar + ar^2 + \cdots + ar^{n-1} = a\frac{1-r^n}{1-r}$ for $r \neq 1$ by finite geometric series.
- $\sum_{j=1}^{n} j = 1 + 2 + \dots + n = \frac{n(n+1)}{2}$ by arithmetic series.
- $\sum_{j=1}^{n} \frac{1}{j} = 1 + \frac{1}{2} + \dots + \frac{1}{n} \approx \log(n) + 0.577$ by harmonic series.
- $\int_0^1 x^{a-1}(1-x)^{b-1}dx = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$ by beta function.
- $\int_0^1 \binom{n}{k} x^k(1-x)^{n-k}dx = \frac{1}{n+1}$ by Bayes' Billiards.
- For $a > 0$, $\Gamma(a) = \int_0^{\infty} x^{a-1}e^{-x}dx$, $\Gamma(a+1) = a\Gamma(a)$ by gamma function.
- For $n \in \mathbb{Z}^+$, $\Gamma(n) = (n-1)!$ by gamma function.

- For $n = \frac{1}{2}$, $\Gamma(\frac{1}{2}) = \sqrt{\pi}$ by gamma function.

Concept Checker 1. Simplify $\sum_{x=0}^{\infty} \frac{e^{-4}6^x}{x!}(0.5)^x$.

Solution

$$\begin{aligned}
\sum_{x=0}^{\infty} \frac{e^{-4}6^x}{x!}(0.5)^x &= e^{-4}\sum_{x=0}^{\infty} \frac{6^x}{x!}(0.5)^x && \text{by factoring} \\
&= e^{-4}\sum_{x=0}^{\infty} \frac{3^x}{x!} && \text{by simplifying} \\
&= e^{-4}e^{3} && \text{by Taylor series} \\
&= e^{-1} && \text{by simplifying}
\end{aligned}$$

3 Probability and Random Variables

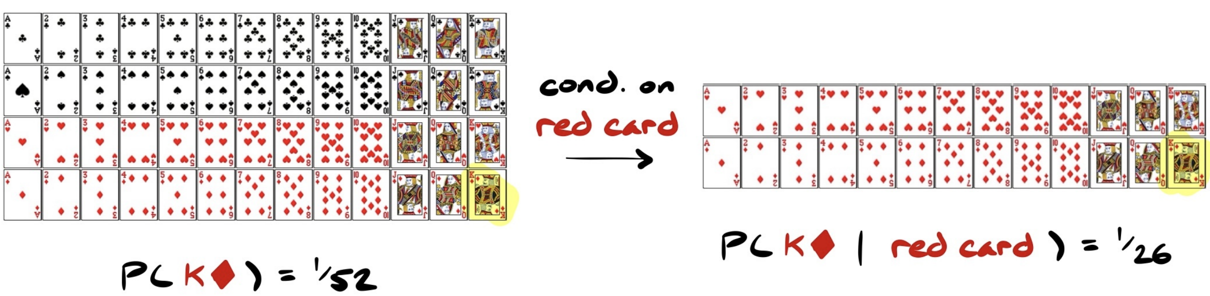

FIGURE 1: A standard deck of 52 playing cards shrunk down to only the 26 red cards.

Definition 1 (Probability)**.** Informally, a measure of how likely an event is, bounded from 0 to 1.[1]

- $P(A) = \frac{\text{\# of outcomes favorable to } A}{\text{total \# of outcomes}}$ by Naive Definition, where all outcomes are equally likely.
- $P(A) \leq P(B)$ if $A \implies B$.
- $P(A \cup B) = P(A) + P(B) - P(A \cap B)$ by Principle of Inclusion-Exclusion.
- $P(A \cap B) = P(A, B) = P(A)P(B \mid A) = P(B)P(A \mid B)$ by Multiplication Rule.
- $P(A) = 1 - P(A^c)$ by Complement Rule, where A^c is the complement of A.
- $P(A) = \sum_{i=1}^{n} P(A \cap B_i) = \sum_{i=1}^{n} P(A \mid B_i)P(B_i)$ by Law of Total Probability, where $B_1, \ldots, B_n$ partition the sample space.
- $P(A \mid B) = \frac{P(A \cap B)}{P(B)} = \frac{P(B|A)P(A)}{P(B)}$ by Bayes' Rule.
- $P((A \cup B)^c) = P(A^c \cap B^c)$, $P((A \cap B)^c) = P(A^c \cup B^c)$ by De Morgan's Laws.

[1]Formally, a probability space consists of a sample space S and a probability function P, which takes an event $A \subseteq S$ as input and returns $P(A)$, a real number between 0 and 1, as output. The function P must satisfy the following axioms: $P(\emptyset) = 0$, $P(S) = 1$, and if $A_1, A_2, \ldots$ are disjoint events, then $P\left(\cup_{i=1}^{\infty} A_i\right) = \sum_{i=1}^{\infty} P(A_i)$. This definition is from *Introduction to Probability* by Joseph K. Blitzstein and Jessica Hwang.

• **Tip**: Conditional probabilities are probabilities. E.g., $P(A \mid B) = 1 - P(A^c \mid B)$ by Complement Rule.

• **Tip**: $B \cap C$ can be thought of as one event. E.g., $P(A \mid B \cap C) = \frac{P(B \cap C \mid A)P(A)}{P(B \cap C)}$ by Bayes' Rule.

• **Tip**: As illustrated in Figure 1, we can think of conditioning as now working with a shrunk-down sample space. Here, we shrink to 26 cards from our original 52.

Definition 2 (Random variable)**.** Informally, an unknown value that will "crystallize" to a number after an experiment.

• For some intuition, we can think of a random variable as analogous to a Mario Kart item box—it is unknown what value it will take on, but through its distribution, we can describe the possible values (along with which ones are more likely).

4 Independence vs. Conditional Independence

Definition 3 (Independence)**.** Informally, for events A, B, knowing B gives no information about A (and vice versa).

• Formally, $A \perp\!\!\!\perp B$ if $P(A \mid B) = P(A)$, assuming $P(B) > 0$.

Definition 4 (Conditional independence)**.** Informally, for events A, B, C, given C, knowing B gives no information about A (and vice versa).

• Formally, $(A \perp\!\!\!\perp B) \mid C$ if $P(A \mid B, C) = P(A \mid C)$, assuming $P(B, C) > 0$ and $P(C) > 0$.

Example 1 (Unconditional independence)**.** I have two fair dice, so the rolls are independent (e.g., if I roll a 6 on the first one, I still don't know what I'll get on the next one). But if we condition on the event the sum is 10, the second roll is deterministic from the first one and thus not conditionally independent.

• Formally, let X_i be the value of my ith roll. $\{X_1 = 6\} \perp\!\!\!\perp \{X_2 = 4\}$, but $(\{X_1 = 6\} \not\perp\!\!\!\perp \{X_2 = 4\}) \mid \{X_1 + X_2 = 10\}$.

Example 2 (Conditional independence)**.** Let p be the probability I make a free throw. If I shoot 9 free throws that all miss, you'd probably assume the 10th shot would miss because the event I miss 9 shots implies p is low. But if we condition on $p = 0.5$, then the 10th shot is conditionally independent of the previous 9; I'm just having an unlucky day.

• Formally, let A be the event I miss the first 9 shots and B be the event I miss the 10th shot. $A \not\perp\!\!\!\perp B$, but $(A \perp\!\!\!\perp B) \mid \{p = 0.5\}$.

5 Expectation

Definition 5 (Expectation)**.** Informally, the weighted average of a random variable.

• $\mathbb{E}[aX + b] = a\mathbb{E}[X] + b$ for constants a, b by linearity.

• $\mathbb{E}[X] = \sum_{\text{supp}(X)} xP(X = x)$ for X discrete by definition.

• $\mathbb{E}[X] = \int_{\text{supp}(X)} xf_X(x)dx$ for X continuous by definition.

• $\mathbb{E}[X] = \mathbb{E}[\mathbb{1}_1 + \cdots + \mathbb{1}_n]$ where $X = \mathbb{1}_1 + \cdots + \mathbb{1}_n$ by substitution.

- $\mathbb{E}[X] = \mathbb{E}[\mathbb{E}[X \mid Y]]$ by Adam's Law.
- $\mathbb{E}[X] = \sum_{i=1}^{n} \mathbb{E}[X \mid A_i]P(A_i)$ by Law of Total Expectation.
- $\mathbb{E}[X^2] = \text{Var}[X] + (\mathbb{E}[X])^2$ by definition of variance.
- $\mathbb{E}[XY] = \mathbb{E}[\mathbb{E}[XY \mid X]] = \mathbb{E}[X \cdot \mathbb{E}[Y \mid X]]$ by taking out what's known.
- $\mathbb{E}[XY] = \text{Cov}[X, Y] + \mathbb{E}[X]\mathbb{E}[Y]$ by definition of covariance.
- $\mathbb{E}[X \mid Y] = \mathbb{E}[\mathbb{E}[X \mid Z, Y] \mid Y]$ by Adam's Law with extra conditioning on Y.
- $\mathbb{E}[X \mid Y = y] = \frac{\mathbb{E}[X \mathbb{1}\{Y=y\}]}{P(Y=y)}$ for Y discrete and $P(Y = y) > 0$ by rearranging LOTE.
- $\mathbb{E}[g(X)] = \sum_{\text{supp}(X)} g(x)P(X = x)$ for X discrete by Law of the Unconscious Statistician.
- $\mathbb{E}[g(X)] = \int_{\text{supp}(X)} g(x)f_X(x)dx$ for X continuous by Law of the Unconscious Statistician.
- $\mathbb{E}[g(X)] \geq g(\mathbb{E}[X])$ for g convex by Jensen's Inequality.
- $\mathbb{E}[g(X)] \leq g(\mathbb{E}[X])$ for g concave by Jensen's Inequality.

Concept Checker 2. Let $X \sim \text{Pois}(\lambda)$. How would you find $\mathbb{E}[X^2]$? Why might it not be a good idea to use LOTUS?

Solution

First, $X \sim \text{Pois}(\lambda)$, so $\mathbb{E}[X] = \text{Var}[X] = \lambda$ as a property of the Poisson. From above, $\mathbb{E}[X^2] = \text{Var}[X] + (\mathbb{E}[X])^2$ by definition of variance, so $\mathbb{E}[X^2] = \lambda + \lambda^2$.
Now, we could've tried using LOTUS (indeed, we're trying to find the expectation of a function of X). But $\mathbb{E}[X^2] = \sum_{x=0}^{\infty} x^2 \frac{e^{-\lambda}\lambda^x}{x!}$ by LOTUS, which does not look fun to solve! In general, for $\mathbb{E}[X^2]$, it's a good idea to try the definition of variance first.

Concept Checker 3. Let X, Y be random variables, with $X \sim \mathcal{N}(1, 1)$, $Y \sim \text{Expo}(\frac{1}{5})$, and $\text{Cov}[X, Y] = 3$. How would you find $\mathbb{E}[XY]$?

Solution

From above, $\mathbb{E}[XY] = \text{Cov}[X, Y] + \mathbb{E}[X]\mathbb{E}[Y]$ by definition of covariance. Thus, $\mathbb{E}[XY] = 3 + (1)(5) = 8$.

Concept Checker 4. Let X, p be random variables. Additionally, let $p \sim \text{Unif}(0, 1)$ and $X \mid p \sim \text{Bern}(p)$. How would you find $\mathbb{E}[X]$?

Solution

$$\begin{aligned}
\mathbb{E}[X] &= \mathbb{E}[\mathbb{E}[X \mid p]] && \text{by Adam's Law} \\
&= \mathbb{E}[p] && \text{since } X \mid p \sim \text{Bern}(p) \\
&= 0.5 && \text{since } p \sim \text{Unif}(0, 1)
\end{aligned}$$

Concept Checker 5. Suppose you want to continue buying lottery tickets until you get a winning ticket. Assume each ticket has a winning probability of $\frac{1}{50}$. Let X be the number of losing tickets purchased (excluding the winning ticket). How would you find $\mathbb{E}[X]$?

Solution

$X \sim \text{Geom}(\frac{1}{50})$ by the story of the Geometric. Thus, $\mathbb{E}[X] = \frac{49/50}{1/50} = 49$ as a property of the Geometric. In general, it's helpful to see if X matches the "story" of a "famous" distribution before doing a bunch of math.

Definition 6 (Conditional expectation)**.** Informally, expectation takes a random variable as input and returns a number as output.

- This is true for unconditional expectation and expectation conditional on an event.
- However, expectation conditional on another random variable is a random variable (specifically, a function of possible crystallizations of that other random variable).
- E.g., let X be someone's yearly salary and Y be someone's daily salary. Then $\mathbb{E}[X \mid Y] = 365Y$, a random variable. Notice $\mathbb{E}[X \mid Y = 1] = 365$, $\mathbb{E}[X \mid Y = 2] = 730$, etc.

Concept Checker 6. What is $\mathbb{E}[X \mid X]$?

Solution

$\mathbb{E}[X \mid X] = X \cdot \mathbb{E}[1] = X \cdot 1 = X$ by taking out what's known. Intuitively, this is a function of possible crystallizations for X; if we're given a value of X, then we expect X to be that value!

Concept Checker 7. What is $\mathbb{E}[X \mid Y]$ if $X \perp\!\!\!\perp Y$?

Solution

$\mathbb{E}[X \mid Y] = \mathbb{E}[X]$ by independence. As an unconditional expectation, this is a number (which can be thought of as a "degenerate" random variable).

6 Variance

Definition 7 (Variance)**.** Informally, a measure of the spread of a random variable, which must be non-negative.

- $\text{Var}[aX + b] = a^2\text{Var}[X]$ for constants a, b by bilinearity.
- $\text{Var}[X] = \mathbb{E}\left[(X - \mathbb{E}[X])^2\right] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$ by definition.
- $\text{Var}[X] = (\text{SD}[X])^2$ by definition of standard deviation.
- $\text{Var}[X] = \mathbb{E}\left[\text{Var}[X \mid Y]\right] + \text{Var}\left[\mathbb{E}[X \mid Y]\right]$ by Eve's Law.
- $\text{Var}[X \mid Y] = \mathbb{E}\left[\text{Var}[X \mid Z, Y] \mid Y\right] + \text{Var}\left[\mathbb{E}[X \mid Z, Y] \mid Y\right]$ by Eve's Law with extra conditioning on Y.
- $\text{Var}[X_1 + \ldots + X_n] = \sum_{i=1}^n \text{Var}[X_i] + 2\sum_{i<j} \text{Cov}[X_i, X_j]$ by bilinearity.

Concept Checker 8. By definition, variance is the expected squared deviation of a random variable from its expectation: $\text{Var}[X] = \mathbb{E}[(X - \mathbb{E}[X])^2]$. Show $\text{Var}[X] = \mathbb{E}[X^2] - (\mathbb{E}[X])^2$.

Solution

First, let $\mu = \mathbb{E}[X]$ to make the notation cleaner.

$$\begin{aligned}
\text{Var}[X] &= \mathbb{E}[(X-\mu)^2] && \text{by definition}\\
&= \mathbb{E}[X^2 - 2X\mu + \mu^2] && \text{by FOIL}\\
&= \mathbb{E}[X^2] - \mathbb{E}[2X\mu] + \mathbb{E}[\mu^2] && \text{by linearity}\\
&= \mathbb{E}[X^2] - 2\mu\mathbb{E}[X] + \mu^2 && \text{by linearity}\\
&= \mathbb{E}[X^2] - 2\mu^2 + \mu^2 && \text{by substitution}\\
&= \mathbb{E}[X^2] - \mu^2 && \text{by simplifying}\\
&= \mathbb{E}[X^2] - (\mathbb{E}[X])^2 && \text{by substitution}
\end{aligned}$$

7 Covariance and Correlation

Definition 8 (Covariance)**.** Informally, a measure of the co-movement of two random variables.

- $\text{Cov}[X,Y] = \mathbb{E}\left[(X - \mathbb{E}[X])(Y - \mathbb{E}[Y])\right] = \mathbb{E}[XY] - \mathbb{E}[X]\mathbb{E}[Y]$ by definition.
- $\text{Cov}[X,Y] = \text{Cov}[Y,X]$.
- $\text{Cov}[X,c] = 0$ for constant c.
- $\text{Cov}[cX+b,Y] = c\text{Cov}[X,Y]$ for constants c, b.
- $\text{Cov}[X,X] = \text{Var}[X]$.
- $\text{Cov}[W+X,Y+Z] = \text{Cov}[W,Y] + \text{Cov}[W,Z] + \text{Cov}[X,Y] + \text{Cov}[X,Z]$.
- $X \perp\!\!\!\perp Y \implies \text{Cov}[X,Y] = 0$, but generally, $\text{Cov}[X,Y] = 0 \not\Longrightarrow X \perp\!\!\!\perp Y$.

Definition 9 (Correlation)**.** Informally, a standardized measure of covariance, bounded from -1 to 1.

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

Concept Checker 9. Let $X_1, \ldots, X_n$ be i.i.d. random variables with $\mathbb{E}[X_i] = \mu$ and $\text{Var}[X_i] = \sigma^2 < \infty$. How can we rewrite $\text{Var}[X_1 + \ldots + X_n]$?

Solution

First, $\text{Var}[X_1 + \ldots + X_n] = \sum_{i=1}^n \text{Var}[X_i] + 2\sum_{i<j}\text{Cov}[X_i, X_j]$ by bilinearity. Next, $\forall i, j$, $X_i \perp\!\!\!\perp X_j$ since they are i.i.d., so $\text{Cov}[X_i, X_j] = 0$. Related, $\forall i, j$, $\text{Var}[X_i] = \text{Var}[X_j]$ since they are i.i.d. Thus, $\text{Var}[X_1 + \ldots + X_n] = \sum_{i=1}^n \sigma^2 + 2\sum_{i<j} 0 = n\sigma^2$.

Concept Checker 10. Now suppose the random variables are identically distributed but not independent. However, they have the same covariance relationships. I.e., $\text{Cov}[X_i, X_j]$ is the same for every pair with $i \neq j$. How can we rewrite $\text{Var}[X_1 + \ldots + X_n]$?

Solution

From before, $\text{Var}[X_1 + \ldots + X_n] = \sum_{i=1}^n \text{Var}[X_i] + 2\sum_{i<j} \text{Cov}[X_i, X_j]$ by bilinearity. We can replace $\text{Var}[X_i]$ with σ^2, but we need to keep $\text{Cov}[X_i, X_j]$ as is. However, since every pair with $i \neq j$ has the same covariance (and there are $\binom{n}{2}$ pairs), we can rewrite the sum. Thus, $\text{Var}[X_1 + \ldots + X_n] = \sum_{i=1}^n \sigma^2 + 2\sum_{i<j} \text{Cov}[X_i, X_j] = n\sigma^2 + 2\binom{n}{2}\text{Cov}[X_i, X_j]$.

8 Distributions

Definition 10 (Distribution)**.** Informally, a way to describe a random variable through its possible crystallizations and their associated probabilities.

• Formally, every random variable X has a CDF: $F(x) = P(X \leq x)$. This is a function of x.

• If X is discrete, it has a PMF: $P(X = x)$. To get the probability X crystallizes to a number a, we evaluate: $P(X = a)$.

• If X is continuous, it has a PDF: $f(x)$. To get the probability X crystallizes within a range $[a, b]$, we integrate: $P(a \leq X \leq b) = \int_a^b f(x)dx = F(b) - F(a)$.

• Informally, $f(x) \xrightarrow{\int_{-\infty}^x dt} F(x)$, $F(x) \xrightarrow{\frac{d}{dx}} f(x)$.

• In general, it's helpful to see if X matches the "story" of a "famous" distribution before doing a bunch of math. See the last page for a list (along with useful properties).

Definition 11 (Universality of the Uniform)**.** Let X be a random variable with CDF F, where F is continuous and strictly increasing on $\text{supp}(X)$. Thus, F^{-1} exists.

• $F(X) \sim \text{Unif}(0, 1)$.

• Let $U \sim \text{Unif}(0, 1)$. $F^{-1}(U)$ is a random variable with CDF F.

Concept Checker 11. Let $Y_i \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$, with both μ and σ^2 known. Through our random number generator, we only have access to i.i.d. $\text{Unif}(0, 1)$ random variables: $U_1, \ldots, U_n$. How can we create a draw of $Y_1, \ldots, Y_n$ in terms of $U_1, \ldots, U_n$?

Solution

Let Φ^{-1} be the inverse CDF for a Standard Normal, which is continuous and strictly increasing on $\mathbb{R}$. By Universality of the Uniform, $\Phi^{-1}(U)$ is a random variable with CDF Φ (i.e., it is Standard Normal). As a property of the Normal, $\sigma\Phi^{-1}(U) + \mu \sim \mathcal{N}(\mu, \sigma^2)$. Thus, let $Y_i = \sigma\Phi^{-1}(U_i) + \mu \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu, \sigma^2)$.

9 Quantiles (Time Permitting, a Sneak Peek into 111!)

Definition 12 (Inverse of a function)**.** For a function $f(x) = y$, the inverse is $f^{-1}(y) = x$ such that $f^{-1}(f(x)) = f(f^{-1}(x)) = x$.

• **Strategy**: Replace $f(x)$ with y, swap x and y, solve for y, and replace y with $f^{-1}(x)$.

Concept Checker 12. Let $f(x) = \log(x) + 5$. What is $f^{-1}(x)$?

Solution

First, replace $f(x)$ with y: $y = \log(x) + 5$. Next, swap x and y: $x = \log(y) + 5$. Then, solve for y: $y = e^{x-5}$. Finally, replace y with $f^{-1}(x)$: $f^{-1}(x) = e^{x-5}$.

Definition 13 (Cumulative distribution function)**.** For a random variable Y, its cumulative distribution function (CDF) is $F(y) = P(Y \leq y)$.

Definition 14 (Quantile function)**.** For a random variable Y, let F be its CDF. The quantile function of Y is $Q(p) = \min\{y : F(y) \geq p\}$, where $Q(p)$ is the p-quantile of the distribution.

• I.e., the p-quantile is the smallest y such that the CDF at y attains at least a value of p.

• If F is continuous and strictly increasing, then F^{-1} exists (as a "true" inverse), so we use $Q(p) = F^{-1}(p)$ such that $P(Y \leq y) = p \iff F(y) = p \iff y = F^{-1}(p) \iff Q(p) = y$. Notice $F : \mathbb{R} \to [0, 1]$ while $F^{-1} : [0, 1] \to \mathbb{R}$.

• If F^{-1} doesn't "truly" exist (e.g., when Y is discrete), we use a "generalized" inverse for the quantile function: $Q(p) = \min\{y : F(y) \geq p\}$ such that $P(Y \leq y) \geq p \iff Q(p) \leq y$.

• E.g., suppose SAT score is distributed $\mathcal{N}(1000, 200^2)$. If we're interested in the 0.5-quantile (i.e., median), then $Q(0.5) = 1000$ since symmetric distributions have mean = median. Notice 50% of the distribution lies to the left of $y = 1000$.

Concept Checker 13. Rewrite the following probabilities as quantiles.

1. Assume F is continuous and strictly increasing. For Z continuous, $P(Z \leq 1.96) = 0.975 \iff$ ____________________

2. Assume $F(3) < F(4)$. For X discrete, $P(X > 4) = 0.05 \iff$ ____________________

Solution

1. $P(Z \leq 1.96) = 0.975 \iff Q(0.975) = 1.96$. In a Standard Normal, 0.975 of the probability lies to the left of 1.96.
2. $P(X > 4) = 0.05 \iff 1 - P(X \leq 4) = 0.05 \iff P(X \leq 4) = 0.95 \iff F(4) = 0.95$. Since every CDF is non-decreasing (i.e., $F(3) \leq F(4) \leq F(5)$, so $F(3) \leq 0.95 \leq F(5)$), the minimum y such that $F(y) \geq 0.95$ is at most 4. Since we assume $F(3) < F(4)$ ($F(4) = 0.95$, so $F(3)$ could be 0.94, e.g.), the minimum y such that $F(y) \geq 0.95$ is 4 (i.e., it cannot be 3 or lower). Thus, $Q(0.95) = 4$. In this distribution, 0.95 of the probability lies to the left of 4.

10 Practice Problems

Problem 1. For $i \in \{1, \ldots, n\}$, let $Y_{i1}, Y_{i2} \overset{\text{i.i.d.}}{\sim} \mathcal{N}(\mu_i, \sigma^2)$, where μ_i is the true expectation for pair i and σ^2 is the true variance shared across all observations. Observations across pairs are independent, so overall, we have $2n$ independent observations: $(Y_{11}, Y_{12}, \ldots, Y_{n1}, Y_{n2})$.[2]

[2]Inspired by Problem 4 in "Stat 111 Homework 4, Spring 2025" by Joseph K. Blitzstein and Neil Shephard.

(a) Find the distribution of $Y_{i1} - Y_{i2}$.

(b) **Challenge**: Find $\text{Var}[\frac{1}{4n}\sum_{i=1}^{n}(Y_{i1} - Y_{i2})^2]$.

Hint: Try to "pattern-match" between the Normal and another known distribution.

Solution

First, $Y_{i1} - Y_{i2} \sim \mathcal{N}(0, 2\sigma^2)$ as a property of the Normal.
Next, let $D_i = Y_{i1} - Y_{i2}$. Let's find $\text{Var}[D_i^2]$.

$$\begin{aligned}
D_i \sim \mathcal{N}(0, 2\sigma^2) &\implies \frac{D_i}{\sigma\sqrt{2}} \sim \mathcal{N}(0,1) && \text{by standardizing} \\
&\implies \frac{D_i^2}{2\sigma^2} \sim \chi_1^2 && \text{by story of Chi-squared} \\
&\implies \text{Var}\left[\frac{D_i^2}{2\sigma^2}\right] = 2 && \text{by property of Chi-squared} \\
&\implies \frac{1}{4\sigma^4}\text{Var}\left[D_i^2\right] = 2 && \text{by bilinearity} \\
&\implies \text{Var}\left[D_i^2\right] = 8\sigma^4 && \text{by algebra}
\end{aligned}$$

With this, we can find $\text{Var}[\frac{1}{4n}\sum_{i=1}^{n}(Y_{i1} - Y_{i2})^2]$.

$$\begin{aligned}
\text{Var}\left[\frac{1}{4n}\sum_{i=1}^{n}(Y_{i1} - Y_{i2})^2\right] &= \text{Var}\left[\frac{1}{4n}\sum_{i=1}^{n} D_i^2\right] && \text{by substituting} \\
&= \frac{1}{16n^2}\left(\sum_{i=1}^{n}\text{Var}[D_i^2] + 2\sum_{i<j}\text{Cov}[D_i^2, D_j^2]\right) && \text{by bilinearity} \\
&= \frac{1}{16n^2}\left(\sum_{i=1}^{n}\text{Var}[D_i^2]\right) && \text{by independence} \\
&= \frac{1}{16n^2}\left(\sum_{i=1}^{n} 8\sigma^4\right) && \text{by substituting} \\
&= \frac{\sigma^4}{2n} && \text{by simplifying}
\end{aligned}$$

Problem 2. Suppose the times of major earthquakes follow a Poisson process of rate λ eruptions per year. A scientist collects data on the major earthquakes in a pre-specified time interval $[0, t]$, where t is fixed and known (in years). Let N be the number of earthquakes in this time interval.[3]

(a) Find the distribution of N.

(b) **Challenge**: Find $\mathbb{E}[\frac{t}{N+1}]$.

Hint: Rewrite the expectation as a series and then "pattern-match" to a famous series.

[3] Inspired by Problem 1 in "Stat 111 Homework 3, Spring 2025" by Joseph K. Blitzstein and Neil Shephard.

Solution

First, $N \sim \text{Pois}(\lambda t)$ since this is a Poisson process.
Next, let's find $\mathbb{E}[\frac{t}{N+1}]$.

$$\begin{aligned}
\mathbb{E}[\frac{t}{N+1}] &= t\mathbb{E}[\frac{1}{N+1}] && \text{by linearity} \\
&= t\sum_{n=0}^{\infty}(\frac{1}{n+1})(\frac{e^{-\lambda t}(\lambda t)^n}{n!}) && \text{by LOTUS} \\
&= t\sum_{n=0}^{\infty}\frac{e^{-\lambda t}(\lambda t)^n}{(n+1)n!} && \text{by simplifying} \\
&= te^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{(n+1)!} && \text{by simplifying}
\end{aligned}$$

Now, $e^x = \sum_{n=0}^{\infty}\frac{x^n}{n!} = 1 + x + \frac{x^2}{2!} + \ldots$ by Taylor series. We want to "pattern-match" to this. We have $\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{(n+1)!} = 1 + \frac{\lambda t}{2!} + \frac{(\lambda t)^2}{3!} + \ldots$, but there's a "disconnect" with the n in the numerator and the $n+1$ in the denominator.
Let's take what we have and multiply it: $(\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{(n+1)!})(\lambda t) = \sum_{n=0}^{\infty}\frac{(\lambda t)^{n+1}}{(n+1)!}$. This is starting to look better! Notice $\sum_{n=0}^{\infty}\frac{(\lambda t)^{n+1}}{(n+1)!} = \lambda t + \frac{(\lambda t)^2}{2!} + \frac{(\lambda t)^3}{3!} + \ldots$, but $e^{\lambda t} = \sum_{n=0}^{\infty}\frac{(\lambda t)^n}{n!} = 1 + \lambda t + \frac{(\lambda t)^2}{2!} + \frac{(\lambda t)^3}{3!} + \ldots$ by Taylor series. Thus, $\sum_{n=0}^{\infty}\frac{(\lambda t)^{n+1}}{(n+1)!} = e^{\lambda t} - 1$.
Putting this all together, $\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{(n+1)!} = \frac{\sum_{n=0}^{\infty}\frac{(\lambda t)^{n+1}}{(n+1)!}}{\lambda t} = \frac{e^{\lambda t}-1}{\lambda t}$.

$$\begin{aligned}
\mathbb{E}[\frac{t}{N+1}] &= te^{-\lambda t}\sum_{n=0}^{\infty}\frac{(\lambda t)^n}{(n+1)!} && \text{from before} \\
&= te^{-\lambda t}(\frac{e^{\lambda t}-1}{\lambda t}) && \text{by substituting} \\
&= \frac{1}{\lambda}(1 - e^{-\lambda t}) && \text{by simplifying}
\end{aligned}$$

Problem 3. The Weibull distribution is a generalization of the Exponential. I.e., for $T \sim \text{Expo}(\lambda)$, $Y = T^{\frac{1}{\gamma}} \sim \text{Weibull}(\lambda, \gamma)$, with γ known and λ unknown. Let $\gamma > 0$. (Different parameterizations of the Weibull exist, so be careful if looking at other sources.)[4]

(a) **Challenge**: Find $\mathbb{E}[Y]$.

Hint: Rewrite the expectation as an integral and then "pattern-match" to a famous integral. It may help to define $x = \lambda t$.

[4] Inspired by Problem 3 in "Stat 111 Homework 2, Spring 2025" by Joseph K. Blitzstein and Neil Shephard.

Solution

$$\begin{aligned} \mathbb{E}[Y] &= \mathbb{E}[T^{\frac{1}{\gamma}}] && \text{by substituting} \\ &= \int_0^\infty t^{\frac{1}{\gamma}} \lambda e^{-\lambda t} dt && \text{by LOTUS} \end{aligned}$$

Let $x = \lambda t$, which implies $t = \frac{1}{\lambda}x$ and $dt = \frac{1}{\lambda}dx$.

$$\begin{aligned} \mathbb{E}[Y] &= \int_0^\infty t^{\frac{1}{\gamma}} \lambda e^{-\lambda t} dt && \text{from before} \\ &= \int_0^\infty (\frac{1}{\lambda}x)^{\frac{1}{\gamma}} \lambda e^{-x} (\frac{1}{\lambda}dx) && \text{by substituting} \\ &= \int_0^\infty (\frac{x}{\lambda})^{\frac{1}{\gamma}} e^{-x} dx && \text{by simplifying} \\ &= \frac{1}{\lambda^{\frac{1}{\gamma}}} \int_0^\infty x^{\frac{1}{\gamma}} e^{-x} dx && \text{by simplifying} \end{aligned}$$

Now, for $a > 0$, $\Gamma(a) = \int_0^\infty x^{a-1} e^{-x} dx$. Thus, $\int_0^\infty x^{\frac{1}{\gamma}} e^{-x} dx = \int_0^\infty x^{(\frac{1}{\gamma}+1)-1} e^{-x} dx = \Gamma(\frac{1}{\gamma} + 1)$.

$$\begin{aligned} \mathbb{E}[Y] &= \frac{1}{\lambda^{\frac{1}{\gamma}}} \int_0^\infty x^{\frac{1}{\gamma}} e^{-x} dx && \text{from before} \\ &= \left(\frac{1}{\lambda^{\frac{1}{\gamma}}}\right) \left(\Gamma(\frac{1}{\gamma} + 1)\right) && \text{by substituting} \end{aligned}$$

Problem 4. The Pareto distribution is a skewed, heavy-tailed distribution used to model phenomena where a large portion of the total is held by a small percentage of the population (e.g., income distribution). Let $X \sim \text{Pareto}(x_m, \alpha)$ with PDF $f(x) = \frac{\alpha x_m^\alpha}{x^{\alpha+1}}$ on the support $x \geq x_m$, where $x_m > 0$, $\alpha > 0$.

(a) Find the CDF: $F(x)$.

(b) Find the quantile function: $Q(p)$.

Solution

First, $F(x) = P(X \leq x) = \int_{-\infty}^x f(t)dt$, where t is the "dummy" variable. Notice the PDF is only defined on the support. Everywhere else, it equals 0.

$$\begin{aligned}
\int_{-\infty}^{x} f(t)dt &= \int_{-\infty}^{x_m} f(t)dt + \int_{x_m}^{x} f(t)dt && \text{by splitting} \\
&= \int_{-\infty}^{x_m} 0dt + \int_{x_m}^{x} \frac{\alpha x_m^\alpha}{t^{\alpha+1}}dt && \text{by substituting} \\
&= \int_{x_m}^{x} \frac{\alpha x_m^\alpha}{t^{\alpha+1}}dt && \text{by simplifying} \\
&= \alpha x_m^\alpha \int_{x_m}^{x} \frac{1}{t^{\alpha+1}}dt && \text{by factoring} \\
&= \alpha x_m^\alpha \int_{x_m}^{x} t^{-\alpha-1}dt && \text{by rewriting} \\
&= \alpha x_m^\alpha [\frac{1}{-\alpha} t^{-\alpha}]|_{x_m}^{x} && \text{by anti-derivative} \\
&= -x_m^\alpha [t^{-\alpha}]|_{x_m}^{x} && \text{by simplifying} \\
&= -x_m^\alpha (x^{-\alpha} - x_m^{-\alpha}) && \text{by evaluating} \\
&= -x_m^\alpha x^{-\alpha} + x_m^\alpha x_m^{-\alpha} && \text{by simplifying} \\
&= 1 - \left(\frac{x_m}{x}\right)^\alpha && \text{by simplifying}
\end{aligned}$$

Thus, $F(x) = 1 - \left(\frac{x_m}{x}\right)^\alpha$. Next, the quantile function is the inverse of the CDF (i.e., $F^{-1}(x)$). To solve for the inverse, let $x = 1 - \left(\frac{x_m}{y}\right)^\alpha$ and solve for y.

$$\begin{aligned}
x = 1 - \left(\frac{x_m}{y}\right)^\alpha &\implies \left(\frac{x_m}{y}\right)^\alpha = 1 - x && \text{by algebra} \\
&\implies \frac{x_m}{y} = (1-x)^{\frac{1}{\alpha}} && \text{by algebra} \\
&\implies y = \frac{x_m}{(1-x)^{\frac{1}{\alpha}}} && \text{by algebra}
\end{aligned}$$

Thus, $F^{-1}(x) = \frac{x_m}{(1-x)^{\frac{1}{\alpha}}}$, so $Q(p) = \frac{x_m}{(1-p)^{\frac{1}{\alpha}}}$.