1. **Definition of Expected Value** (Blitzstein 4.1-4.3, Rice 4.1)

a) Give the definition of the expected value for discrete and for continuous random variables and explain why expectations are not always defined. Use the $t_{(1)}$ (Cauchy) as an example, and explain why we can't just say the mean is 0 due to symmetry.

Discrete Random Variable

$$E(x) = \sum_{i=1}^{\infty} \underbrace{x_i}_{\text{value}} \cdot \underbrace{P(X = x_i)}_{\text{pmf at } x}$$

→ multiply each x_i by it's own probability $P(x = x_i)$ and take the sum of all of these values

Continuous random variable

$$E(x) = \int_{-\infty}^{\infty} x \cdot f_x(x)\, dx$$

(sum over the density function rather than discrete probabilities)

Q: Why are expected values not always defined?

Generally: when the integral or sum diverges

Example: Let's look at the reciprocal gamma density

$X \sim \text{InvGamma}(\alpha, 1)$

$$f(x) = \frac{1}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-1/x} \qquad \alpha = 1/2$$

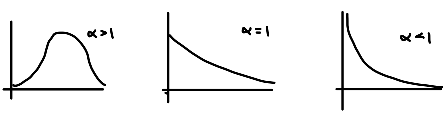
$$E(x) = \int_0^{\infty} x \cdot f(x)\, dx = \int_0^{\infty} \frac{x}{\Gamma(\alpha)} x^{-(\alpha+1)} e^{-1/x}\, dx$$

$x \cdot x^{-(\alpha+1)} = x^{-\alpha}$

$$= \int_0^{\infty} \frac{1}{\Gamma(\alpha)} x^{-\alpha} e^{-1/x}\, dx$$

$$= \frac{1}{\Gamma(\alpha)} \int_0^{\infty} x^{-(\alpha-1+1)} e^{-1/x}\, dx$$

$$= \frac{\Gamma(\alpha-1)}{\Gamma(\alpha)},$$ therefore, α must be greater than one to avoid getting $\emptyset$ in the numerato

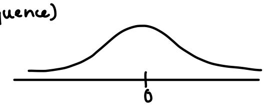


* $\alpha \leq 1$ has unclear $E(x)$

→ too much probability close to $\emptyset$

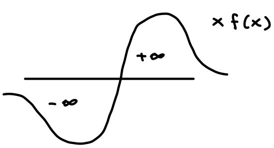
Example: $t_{(1)}$ (cauchy sequence)

$$f(x) = \frac{1}{\pi(1+x^2)}$$



symmetric about zero

$E(x)$ defined iff $\int_{-\infty}^{\infty} |x| f(x)\, dx < \infty$



This visualization creates the false illusion that the integrals of these areas will cancel one another out and make $E(x) = 0$.

$$\int_{-\infty}^{\infty} |x| f(x)\, dx = \frac{2}{\pi} \int_0^{\infty} x f(x)\, dx = \frac{2}{\pi} \int_0^{\infty} \frac{x}{1+x^2}\, dx = \frac{1}{\pi}\left[\log(1+x^2)\right]_0^{\infty} = \infty$$

If we cancel $+\infty - \infty = 0$ BUT then we can make $E(x) = a$ (for any a) and say the $-\infty$ and $+\infty$ on either side will cancel out.

b)* Show that, for a discrete random variable X that takes only non-negative integers, $E(X) = \sum_{k=0}^{\infty} P(X > k)$. Hint: re-express k as a sum from $j = 1$ to k of 1, then reverse the order of summation. Use this to find the expected value of a Geometric(θ) random variable that counts the trials until the first success (compare to using the definition to find $E(X)$).

WTS: $E[X] = \sum_{k=0}^{\infty} P(X > k)$

definition: $E[X] = \sum_{k=0}^{\infty} k \cdot P(X = k)$

$$= \sum_{k=1}^{\infty} \left(\sum_{j=1}^{k} \cdot 1 \right) P(x=k)$$

$$= \sum_{j=1}^{\infty} \sum_{k=j}^{\infty} P(x=k)$$

$$= \sum_{j=1}^{\infty} P(x \geq j)$$

$$= \sum_{j=0}^{\infty} P(x > j)$$

Geometric $X = \#$ of trials at time of first success

iid Bernoulli (p) trials

$P(X = x) = (1-p)^{x-1} p \qquad x = 1, 2, \ldots$

$$E[X] = \sum_{x=0}^{\infty} x (1-p)^{x-1} p$$

$$= p \sum_{x=0}^{\infty} -\frac{d}{dp} (1-p)^{x}$$

$$= p \frac{d}{dp} \sum_{x=0}^{\infty} -(1-p)^{x}$$

$$= p \frac{d}{dp} \left(-\frac{1}{p} \right) = \frac{1}{p}$$

Back to theorem $\rightarrow$ $E[X] = \sum_{k=0}^{\infty} P(X > k)$

Geometric: $P(X > x) = (1-p)^{x}$ *first X trials are all failures

Theorem: $E(X) = \sum_{k=0}^{\infty} P(X > k) = \sum_{k=0} (1-p)^{k} = \frac{1}{p}$

* Geometric Series

Extra notes:

Showing a geometric series converges:

$$\sum_{k=0}^{\infty} \phi^{k} = c$$

$\phi \sum_{k=0}^{\infty} \phi^{k} = c\phi = c - 1 \longrightarrow c(1-\phi) = 1$

$c = \frac{1}{1-\phi}$

$$= \sum_{k=1}^{\infty} \phi^{k} = c - \phi^{0}$$

c) State the results for the linearity of expectation. Review how Blitzstein argues, without using joint distributions, that $E(X+Y) = E(X)+E(Y)$, even if X and Y are not independent. Give examples using extremely correlated variables (e.g., $X_1 = X_2$ and $X_1 = -X_2$), Show that if $E(X_i) = \mu$ for $i = 1, \ldots, n$, then $E(\bar{X}) = \mu$.

Linearity of expectation:

For any random variables X, Y and any constant c,

$$E(X+Y) = E(X) + E(Y)$$
$$E(cX) = cE(X)$$

Blitzstein: $E(X+Y) = E(X) + E(Y)$ even if X and Y are not independent

Consider the extreme case where X always equals Y.

Then, $X + Y = 2X$

and both sides of $E(X+Y) = E(X) + E(Y)$

$$E[X+X] = E[X] + E[X]$$
$$= E[2X] = E[2X]$$
$$= 2E(X) = 2E(X)$$

So linearity still holds even in the most extreme case of <u>dependence</u>

Example using extremely correlated samples:

consider two random variables:

case 1: $X_1 = X_2$

case 2: $X_1 = -X_2$

* in both cases, X_1 and X_2 are highly correlated

linearity of expectation still holds

case 1: $E(X_1 + X_2) = E(2X_1) = 2E(X_1) = E(X_1) + E(X_2)$

case 2: $E(X_1 + X_2) = E(X_1 - X_1) = 0 = E(X_1) - E(X_2)$

In both of these cases, $E(X_1 + X_2) = E(X_1) + E(X_2)$

→ linearity holds

WTS: If $E(X_i) = \mu$ for all $u = 1, \ldots, n$, then, $E[\bar{X}] = \mu$

Sample mean $\bar{X}$ is defined as:

$$\bar{X} = \frac{1}{n} \sum_{i=1}^{n} X_i$$

By the linearity of expectation:

$$E(\bar{X}) = \frac{1}{n} \sum_{i=1}^{n} E(X_i) = \frac{1}{n} \cdot n\mu = \mu$$

Thus, $E(\bar{X}) = \mu$

d) Prove Markov's and Chebyshev's inequalities, and show how these imply the weak law of large numbers. Use the fact that, for $X_1, \ldots, X_n$ iid, $\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$.

Markov's Inequality:

For any r.v. X and constant $a > 0$,

$$P(|X| \geq a) \leq \frac{E|X|}{a}$$

we can cancel this given it is not in the desired bounds

proof: $E(X) = \int_{-\infty}^{\infty} x f(x)\,dx = \int_{x<a} x f(x)\,dx + \int_{x \geq a} x f(x)\,dx$

$$E(X) \geq \int_{x \geq a} x f(x)\,dx \geq \int_{x \geq a} a f(x)\,dx$$

$$E(X) = a P(X \geq a)$$

$$\frac{E(X)}{a} \geq P(X \geq a)$$

Intuitive Interpretation:

set $a = 2E(X)$

Markov's Inequality says: $P(X \geq 2E(X)) \leq 1/2$

i.e. it is impossible for more than half the population to make at least twice the income

(if over half the population were earning at least twice the average income, the income would be higher)

Chebyshev's Inequality:

Let X have mean μ and variance σ^2

Then for any $a > 0$, $P(|X - \mu| \geq a) \leq \frac{\text{Var}(X)}{a^2}$

* The idea for proving chebyshev from markov was to square $|X-\mu|$ and then apply markov.

proof: By Markov's Inequality,

$$P(|X-\mu| \geq a) = P((X-\mu)^2 \geq a^2) \overset{\text{Markov}}{\underset{*}{\leq}} \frac{E(X-\mu)^2}{a^2} = \frac{\text{Var}(X)}{a^2}$$

Weak law of large numbers: $X_1, X_2, \ldots, X_n$ are iid random variables

For all $\varepsilon > 0$, $P(|\bar{X}_n - \mu| > \varepsilon) \to 0$ as $n \to \infty$

(this form of convergence is called convergence in probability)

* $\varepsilon > 0$ is an arbitrary positive number

$E(\bar{X}) = \mu$

$\text{Var}(\bar{X}) = \frac{\sigma^2}{n}$

By Chebyshev's Inequality,

$$P(|\bar{X}_n - \mu| > \varepsilon) \leq \frac{\text{Var}(\bar{X}_n)}{\varepsilon^2} = \frac{\sigma^2}{n\varepsilon^2} \to \emptyset$$

Weak law of large numbers: the probability that the sample mean deviates from the true mean by more than ε becomes arbitrarily small as the sample size increases, implying that the sample mean $\bar{X}_n$ converges to the population mean μ in probability.