Stat 111 Spring 2025 Week 4: Expected Value
Expected Value

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.

   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 $E(X)$).

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

   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, $\mathrm{Var}(\bar{X}) = \frac{\sigma^2}{n}$.

2. **Expectation for Transformations** (Blitzstein 4.5, Rice 4.1-4.2, 4.6,)

   a) Point out that, for a non-linear function $g(x)$, it is unusual for $E(g(X))$ to equate to $g(E(X))$. Give justification for Jensen's inequality and state its implications for expected values. For example, show that $E(1/X) \geq 1/E(X)$, and that $E(\log(X)) \leq \log(E(X)) \leq E(X) - 1$.

   b) State Theorem A of 4.1 in Rice and explain why this is called the law of the unconscious statistician (see also Blitzstein 4.5). Explain how LOTUS is used when we find $\mathrm{Var}(X) = E((X - E(X))^2) = E(X^2) - E(X)^2$. Describe what would be involved to find a variance without using LOTUS (e.g., for a Gamma variable).

   c) State the multivariate version of LOTUS and show this implies $E(X+Y) = E(X)+E(Y)$, even if X and Y are not independent. Describe what would be involved to find a covariance without using LOTUS

   d) Show that, for uncorrelated random variables, $\mathrm{Var}(X + Y) = \mathrm{Var}(X) + \mathrm{Var}(Y)$, but that this is not true in general. Give examples using extremely correlated variables (e.g., $X_1 = X_2$ and $X_1 = -X_2$).

   e) Show how to find approximations to the mean and variance of a transformation of a random variable using Taylor's approximation. As an example, find approximations to the mean and variance of $Y = \log(X)$, for $X \sim \mathrm{Gamma}(\alpha, \lambda)$.

3. **Moment Generating Functions**

   a) Define the MGF $M_x(t)$ for a random variable X and give conditions for it to exist. Show that $Mx(0) = 1$ and that the kth derivative of $M_x(t)$ at $t = 0$ is $E(X^k)$, the kth moment of the distribution (see, e.g., Blitzstein 6.4). Describe the general uniqueness property of the MGF and explain why it is suggested by the moment generation property, and by the connection between the MGF and the invertible Characteristic function.

   b) Find the MGF for $Z \sim N(0, 1)$ and for $X = \mu + \sigma Z \sim N(\mu, \sigma^2)$. Give the general formula for the MGF of a linear transformation of a variable.

c) Show that for independent random variables $X_1, ..., X_n$ with MGF's $M_{x_i}(t)$, the MGF for $Y = \sum X_i$ is $M_y(t) = \prod M_{x_i}(t)$. Use this result to show that the sum of independent Normal variables is also Normal, and that for iid Normal variables, $\bar{X} \sim N(\mu, \frac{\sigma^2}{n})$.

d) State the central limit theorem, and sketch the proof for distributions that have an MGF. Explain why the CLT is not relevant in situations where we assume $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$.

4. **Multivariate MGFs and Sample Statistics** (Rice 4.6, 6.1-6.3)

a) Define the bivariate MGF: $M_{xy}(s,t) = E(e^{sx+ty})$. Show that you get the marginal MGF's by evaluating with $t = 0$ or $s = 0$.

b) Show that if X and Y have a joint MGF, then X and Y are independent if and only if $M_{xy}(t,s) = M_x(t)M_y(s)$.

c) For $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$, show that the average $\bar{X}$ is independent of the sample variance $s^2 = \sum \frac{(X_i - \bar{X})^2}{n-1}$.

d) For $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$, show that the distribution of s^2 is Gamma$(\frac{n-1}{2}, \frac{n-1}{2\sigma^2})$, so that $\frac{(n-1)s^2}{\sigma^2} \sim \chi^2_{(n-1)}$.

e) For $X_1, \ldots, X_n \overset{\text{i.i.d.}}{\sim} N(\mu, \sigma^2)$, show that the distribution of $T = \frac{\bar{X} - \mu}{s/\sqrt{n}}$ is $t_{(n-1)}$.

5. **Conditional Expectation** (Blitzstein 9, Rice 4.4)

a) Distinguish between the conditional expectation given an event, and the conditional expectation given a random variable. As an example, for a Poisson process with rate λ events per unit time, let X be the time of the first event and let Y be the time of the second event. Find $E(X|Y = y)$ and find the distribution of $E(X|Y)$.

b) Define the laws of total expectation and variance. Sketch the proofs of Adam's and Eve's laws (and explain why they have that name).

c)* Use Adam's and Eve's laws to find the mean and variance of a Negative Binomial variable X, where $X|\theta \sim$ Pois(θ), with $\theta \sim$ Gamma$(\alpha, \alpha/\mu)$. Compare E(X) and Var(X) to the Poisson mean and variance, and show how Negative Binomial converges to Poisson as $\alpha \to \infty$.

d) Suppose $\theta \sim N(\mu, A)$ and $Y|\theta \sim N(\theta, V)$. Find the unconditional mean and variance of Y. Represent θ and Y in terms of two independent standard Normal variables Z_1 and Z_2, and use this to show that Y is Normal with this mean and variance.

e) Explain how the model in part d relates to Bayesian inference for the mean of a Normal distribution. What is the posterior distribution of the parameter θ for a given data value $Y = y$? Show the posterior mean is a weighted average of the unbiased estimate Y and the prior mean μ, and that the posterior information (reciprocal variance) is $\frac{1}{V} + \frac{1}{A}$, the sum of the information from the data value Y and the prior information.