Week 5 Stats 25-26

Srikanth Pai, Asst. Prof, MSE.

September 6, 2025

By now we have realised a pattern: Replace summation of pmfs in discrete rv expression with integration of pdfs to obtain formulae for continuous rv expression. CDF is a **fundamental object** in probability theory. So its definition is common!

Discrete Random Variables	**Continuous Random Variables**
Joint CDF: $F_{X,Y}(x,y) = P(X \leq x, Y \leq y)$ (Same definition for both discrete and continuous)	
Marginal CDF from Joint CDF: $F_X(x) = F_{X,Y}(x,\infty) = \lim_{y\to\infty} F_{X,Y}(x,y)$ (Same formula for both discrete and continuous)	
Marginal PMF/PDF: $p_X(x) = \sum_y p_{X,Y}(x,y)$	$f_X(x) = \int_{-\infty}^{\infty} f_{X,Y}(x,y)\, dy$
PMF/PDF from Joint CDF: $p_{X,Y}(x,y) = F_{X,Y}(x,y) - F_{X,Y}(x^-,y)$ $-F_{X,Y}(x,y^-) + F_{X,Y}(x^-,y^-)$	$f_{X,Y}(x,y) = \frac{\partial^2 F_{X,Y}(x,y)}{\partial x \partial y}$
Joint CDF from PMF/PDF: $F_{X,Y}(x,y) = \sum_{u\leq x}\sum_{v\leq y} p_{X,Y}(u,v)$	$F_{X,Y}(x,y) = \int_{-\infty}^{x}\int_{-\infty}^{y} f_{X,Y}(u,v)\, dv\, du$
Independence: $p_{X,Y}(x,y) = p_X(x)\cdot p_Y(y)$	$f_{X,Y}(x,y) = f_X(x)\cdot f_Y(y)$

1. (a) Prove $Cov(X+Y,Z) = Cov(X,Z) + Cov(Y,Z)$.
   (b) Prove $Var(X+Y) = Var(X) + Var(Y) + 2Cov(X,Y)$.

2. Use the above table and solve the following problems from Larsen and Marx: Problems 3.7.1, 3.7.5, 3.7.10, 3.7.19f, 3.7.20, 3.7.27, 3.7.30, 3.7.37, 3.7.41, 3.7.51.

3. In class we derived the normal distribution density by approximating the binomial pmf near the mean. Assuming notations followed in class, show using Taylor series that

$$\ln\sqrt{2\pi N} - \ln\sqrt{2\pi(Np+\rho)} - \ln\sqrt{2\pi(N(1-p)-\rho)} = -\ln\sqrt{2\pi Np(1-p)} + O\left(\frac{\rho}{N}\right).$$

4. We say that a random variable $X \sim \text{Geometric}(p)$ if it takes natural number values and the pmf is

$$p_X(k) = p(1-p)^{k-1} \qquad k \in \mathbb{N}.$$

   (a) Compute the cdf $F_X(x)$ for all real x.
   (b) A factory machine is inspected n times per day, at equal intervals of length $1/n$ days. The probability that it fails during any such interval, given it was working at the start, is approximately λ/n, where $\lambda > 0$ is a fixed constant.

   Let X_n be the number of inspections until the first failure is detected.

      i. Show that X_n follows a geometric distribution with parameter λ/n.
      ii. Let $T_n = X_n/n$ be the number of days (possibly fractional) until failure detection. For $t \geq 0$, compute

$$\lim_{n\to\infty} \Pr(T_n \leq t).$$

      iii. Identify the limiting distribution and explain its meaning in this setting.

5. (V. Imp) A continuous random variable X is normally distributed, i.e. $X \sim \mathcal{N}(\mu, \sigma^2)$ if the pdf is

$$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right) \qquad x \in \mathbb{R}$$

   Assume the following fact proved in a calculus course:

$$\int_{-\infty}^{\infty} e^{-\alpha z^2}\, dz = \sqrt{\frac{\pi}{\alpha}} \quad \text{if } \alpha > 0.$$

   The cdf of the normal distribution does not have a closed form expression.

   (a) Prove the following:
      i. $\mathbb{E}(X) = \mu$
      ii. $\mathbb{E}((X-\mu)^2) = \sigma^2$. [i.e. the variance is σ^2.]
   (b) If $X \sim \mathcal{N}(\mu, \sigma^2)$, and if a new random variable is defined as $Z = \dfrac{X-\mu}{\sigma}$. Calculate the pdf of Z. [In Stats, this will become the famous Z-score!]

6. (Optional for Internal 1, Connection to Linear Algebra class) If you are curious how to prove Cauchy Schwarz inequality, then here is an outline: Let X, Y be two random variables.

   (a) Let $f(t) = \mathbb{E}(X - tY)^2$ be a function defined on all reals. Observe that f is quadratic expression in t.

   (b) Since $f(t) \geq 0$, the discriminant of the quadratic f must be non-positive.

   (c) Expand the square and apply the linearity of expectation in definition of $f(t)$ and set discriminant ≤ 0 to obtain Cauchy-Schwarz.

   (d) Rearrange to show that correlation is between 1 and -1.

   Note that since correlation is in the range $[-1, 1]$, we can find an angle θ so that $\cos\theta = \text{Corr}(X, Y)$.

   In other words, we can construct a vector space V of zero-mean random variables with finite variance and define an inner-product between two random variables as

$$< X, Y >:= Cov(X, Y) = \mathbb{E}(XY).$$

   The length of the random variable X is $\sqrt{Var(X)}$. So we have

$$Cov(X, Y) =< X, Y >= ||X|| \cdot ||Y|| \cos\theta = \sqrt{Var(X)}\sqrt{Var(Y)}\text{Corr}(X, Y)$$

   [**Poorna Ma'am will teach inner products eventually.**]

   COOL FACT: Can you see that Problem 1b becomes "cosine-law" from school in this geometry?!!

   Application to Econ and Finance: A generalisation of this vector space is the fundamental space where all Econometric theory can be well understood. The theory of asset pricing is grounded in a vector space of a similar type where vectors are random payoffs. Time series analysis is about paths in the in the econometric vector space (also called a stochastic process).