MATH 320: Test 2 Study Guide

Lecture 7 – Random Variables (2.1 and 3.1)

Random variables

- Definition: Function from a sample space S into real numbers.
- Range of a RV: The set of possible values of X, $\mathcal{X} = \{x : X(s) = x, s \in S\}$
- RV X is discrete $\iff$ $\mathcal{X}$ is a finite or countable set $\iff$ $F_X(x)$ is a step function of x.
- RV X is continuous $\iff$ $\mathcal{X}$ is an interval (or union of intervals) on the real number line $\iff$ $F_X(x)$ is a continuous function of x.

Lecture 8 – Distribution Functions (2.1 and 3.1)

Calculating probabilities

- Definition: The probability mass function (pmf) of a discrete random variable X is given by

  $f_X(x) = P(X = x), \quad \text{for all } x$
- Definition: A probability density function (pdf) is a continuous random variable X is a real-valued function that can be used to find probabilities using

  $$P(a \leq X \leq b) = \int_a^b f(x)\,\mathrm{d}x$$

  $$\text{For } a \in \mathcal{X},\; P(X = a) = \int_a^a f(x)\,\mathrm{d}x = 0 \quad \Longrightarrow \quad \text{For } (a, b) \in \mathcal{X}, \quad P(a < X < b) = P(a \leq X \leq b)$$

Valid pmfs and pdfs

- Theorem: A function $f_X(x)$ is a pdf (or pmf) of a random variable X if and only if
  (a) $f_X(x) \geq 0$ for all x.
  (b) $\sum_x f_X(x) = 1$ (pmf) or $\int_{-\infty}^{\infty} f_X(x)\,\mathrm{d}x = 1$ (pdf).

Cumulative distribution function (cdf)

- Definition: $F_X(x) = P_X(X \leq x), \quad -\infty < x < \infty$
- Properties of cdfs:
  1. The cdf is defined for $-\infty < x < \infty$ always.
  2. The range of every cdf is $0 \leq F(x) \leq 1 \iff$ Limits: $\lim_{x \to -\infty} F(x) = 0$ and $\lim_{x \to \infty} F(x) = 1$
  3. $F_X(x)$ is a non-decreasing function.

4. If X is discrete $\to F(x)$ is a right continuous step function.

   If X is continuous $\to F(x)$ is a continuous function.

- Relationship between continuous cdf and pdf

  $F'(x) = f(x)$, or equivalently $\frac{\mathrm{d}}{\mathrm{d}x} F_X(x) = f_X(x)$

- Alternate definition of pdf:

  The pdf of a continuous random variable X as the function that satisfies $F_X(x) = \displaystyle\int_{-\infty}^{x} f(t)\,\mathrm{d}t \quad \text{for all } x.$

Finding probabilities using the cdf

- Cdf always gives a left probability.
- If X is discrete, $F(a) = P(X \leq a) = \displaystyle\sum_{x \leq a} f(x)$

  "Complement of cdf": $1 - F(x) = 1 - P(X \leq x) = P(X > x)$

  Interval probabilities: $P(a < X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$

- If X is continuous: $F_X(x) = \displaystyle\int_{-\infty}^{x} f(t)\,\mathrm{d}t$

  For a specific value of $x = a$, we find probability with: $F(a) = \displaystyle\int_{-\infty}^{a} f(x)\,\mathrm{d}x$

  Complement of cdf: $1 - F(a) = 1 - P(X \leq a) = 1 - F(a)$

  Interval probabilities: $P(a \leq X \leq b) = P(X \leq b) - P(X \leq a) = F(b) - F(a)$

Lecture 9 – Summary Measures (2.2, 2.3 and 3.1)

Expected value

- Definition:

  If X is discrete $\to \mu = E(X) = \sum x\, f(x)$

  If X is continuous $\to \mu = E(X) = \displaystyle\int_{-\infty}^{\infty} x\, f(x)\,\mathrm{d}x$

Expected value of a function of a random variable

- If $Y = aX + b \to E(Y) = E(aX + b) = aE(X) + b$
- If X is discrete:

  (Used in the derivation of the above identity) If $Y = aX + b \to f_Y(y) = f_Y(ax + b) = f_X(x)$

  In general, if $Y = g(X) \to E(Y) = \displaystyle\sum_{y} y\, f(y) = E[g(X)] = \sum_{x} g(x)\, f(x)$

- If X is continuous $\rightarrow E(Y) = \int_{-\infty}^{\infty} y\, f(y)\, \mathrm{d}y = E[g(X)] = \int_{-\infty}^{\infty} g(x)\, f(x)\, \mathrm{d}x$
- Linear / Distributive property of expectation:

  $$E\left[\sum_{i=1}^{k} c_i g_i(X)\right] = \sum_{i=1}^{k} c_i E[g_i(X)]$$

Variance and standard deviation

- Variance definitions:

$$V(X) = \begin{cases} & \underline{\text{In general}} & & \underline{\text{Discrete}} & \underline{\text{Continuous}} \\ 0) & \sigma^2 & & & \\ 1) & E[(X-\mu)^2] & \rightarrow & \sum (x-\mu)^2\, f(x) & \int_{-\infty}^{\infty} (x-\mu)^2\, f(x)\, \mathrm{d}x \\ 2) & E(X^2) - \mu^2 & \rightarrow & \sum x^2\, f(x) - \left[\sum x\, f(x)\right]^2 & \int_{-\infty}^{\infty} x^2\, f(x)\, \mathrm{d}x - \left[\int_{-\infty}^{\infty} x\, f(x)\, \mathrm{d}x\right]^2 \end{cases}$$

- Using variance definition 2) $\Rightarrow E(X^2) = V(X) + [E(X)]^2$
- Standard deviation definition: $\sigma = SD(X) = \sqrt{V(X)}$

Variance and standard deviation of $Y = aX + b$

- If $Y = aX + b \rightarrow \sigma_Y^2 = V(Y) = V(aX+b) = a^2 V(X) = a^2 \sigma_X^2$

  $\Rightarrow \sigma_Y = SD(Y) = SD(aX+b) = |a|\, SD(X) = |a|\, \sigma_X$

Mode

- Definition: Mode is the x value which maximizes the distribution function $f(x)$.

Median and Percentiles

- Median m of a continuous random variable X is the solution to: $F(m) = P(X \leq m) = 0.5$.
- Percentile: For $0 \leq p \leq 1$, the $100p^{th}$ percentile of X is the number x_p defined by $F(x_p) = p$.
- $IQR = Q_3 - Q_1$.