WMS: 3.3

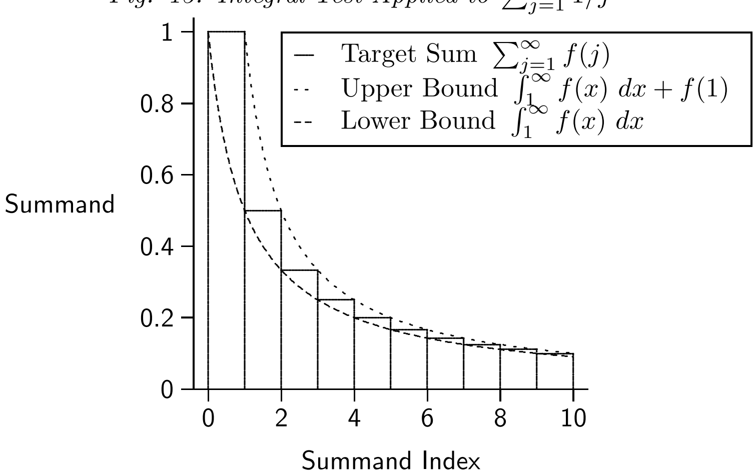
D. The Expectation
1. The expectation represents a mean or average value.
   a. Suppose random variable X
      i. Set of possible values $\mathcal{X}$
      ii. probability function $p_X(x)$
   b. $\mathsf{E}(X) = \sum_{x\in\mathcal{X}} x p_X(x)$.
   c. Operationalize expressing $\mathcal{X}$ as list indexed by integers, and do traditional infinite sum.
      i. Express $\mathcal{X}$ as $\{x_1, x_2, \ldots\}$.
      ii. $\mathsf{E}(X) = \sum_{j=1}^{\infty} x_j p_X(x_j)$.
   d. Defines a typical value
      i. Advantage: explicitly and uniquely defined.
         - Explicit in that I gave you a formula above that returns a number
         - Unique: Does it depend on how we chose to express $\mathcal{X}$?
      ii. Disadvantage: Sometimes isn't defined.
2. Examples:
   a. Single Die
      i. $\mathcal{X} = \{1, \ldots, 6\}$
      ii. $p_X(x) = 1/6$ for all $x \in \mathcal{X}$.
      iii. $\mathsf{E}(X) = 1\times\frac{1}{6}+2\times\frac{1}{6}+3\times\frac{1}{6}+4\times\frac{1}{6}+5\times\frac{1}{6}+6\times\frac{1}{6} = 21/6 = 3.5$
   b. *Bernoulli trial*
      i. Variable takes on value 1 with some probability $\pi \in [0, 1]$
      ii. Variable is zero otherwise.
      iii. $\mathcal{X} = \{0, 1\}$

      iv. $\mathsf{E}(X) = 0 \times (1 - \pi) + 1 \times \pi = \pi$.
3. Define expectation only when expectation of absolute value is finite.
   a. Note $\mathsf{E}(X) = \sum_{x\in\mathcal{X},x<0} x p_X(x) + \sum_{x\in\mathcal{X},x>0} x p_X(x)$.
   b. Problem: if $\sum_{x\in\mathcal{X}} |x|\, p_X(x) = \infty$ then either $\sum_{x\in\mathcal{X},x<0}(-x)p_X(x) = \infty$ or $\sum_{x\in\mathcal{X},x>0} x p_X(x) = \infty$ or both.
      i. In the last case, $\infty - \infty$ is ambiguous.
      ii. In the $\infty - \infty$ case, generally, one find two different expressions $\mathcal{X} = \{x_1, x_2, \ldots\}$ and $\mathcal{X} = \{y_1, y_2, \ldots\}$ so that $\sum_{i=1}^{n} x_j p_X(x_j)$ and $\sum_{i=1}^{n} y_j p_X(y_j)$ do not converge to the same limit.
   c. Don't define expectation if $\sum_{x\in\mathcal{X}} |x|\, p_X(x) = \infty$.
4. A counterexample for which the expectation doesn't exist.
   a. Suppose $\mathsf{P}(X = j) = j^{-2}/c$ for $j = 1, 2, \cdots$.
   b. To make these probabilities sum to 1, $c = \sum_{j=1}^{\infty} j^{-2}$.
      i. Integral test shows that c finite:

$$\int_1^\infty \frac{dx}{x^2} = \lim_{a\to\infty} \int_1^a \frac{dx}{x^2}\, dx = \lim_{a\to\infty} \left.\frac{-1}{x}\right|_1^a = \lim_{a\to\infty}(1 - \frac{1}{a}) = 1 < \infty.$$

      ii. Euler showed that $c = \pi^2/6$, but we won't need that.
   c. However, $\mathsf{E}(X) = \sum_1^\infty j^{-1}/c = \infty$.
      i. Integral test to see sum infinite: $\int_1^\infty (1/x)\ dx = \lim_{a\to\infty} \int_1^a (1/x)\ dx = \lim_{a\to\infty} \ln(x)|_1^a = \lim_{a\to\infty} \ln(a) = \infty$
      ii. See Fig. 13.

Fig. 13: Integral Test Applied to $\sum_{j=1}^{\infty} 1/j$



5. The logarithm to be used in class is the natural log.
   a. Here and everywhere else it appears in class, $\ln(x)$ is the natural log function.
      i. Satisfies $e^{\ln(x)} = \exp(\ln(x)) = x$
   b. There are other alternative log definitions.
      i. common log $\log_{10}(x)$ satisfying $10^{\log_{10}(x)} = x$.
         - Called "common" because it was a tool for performing multiplications before the advent of floating-point portable calculators.
         - Also a device for measuring ship's speed in knots
      ii. Base-2 log $\log_2(x)$ satisfying $2^{\log_2(x)} = x$.
6. Expectation of a transformation of a random variable

   a. For now, restrict attention to discrete random variables.
   b. First construct probability function of a transformation of a random variable $r(X)$.
      i. Suppose that $Y = r(X)$ for some function r .
      ii. Want $p_Y(y)$.
      iii. Let $r^{-1}(\{y\}) = \{x | r(x) = y\}$ be the set of values for X giving a Y value of y .
         - Note that $\{s|Y(s) = y\} = \cup_{x\in r^{-1}(\{y\})}\{s|X(s) = x\}$.
         - Note that if $x_1 \neq x_2$, then $\{s|X(s) = x_1\} \cap \{s|X(s) = x_2\} = \emptyset$.
         - Then

$$p_Y(y) = \mathsf{P}(r(X) = y) = \sum_{x\in r^{-1}(\{y\})} \mathsf{P}(X = x) = \sum_{x\in r^{-1}(\{y\})} p_X(x).$$

   c. Expectation $\mathsf{E}(r(X))$ is defined using original definition for new variable.
      i. Make new random variable $Y = r(X)$.
      ii. Determine range of possible values $\mathcal{Y}$.
      iii. Calculate probability function $p_Y(y)$
      iv. Report $\sum_{y\in\mathcal{Y}} y p_Y(y)$.
      v. Note $\mathcal{X} = \cup_{y\in\mathcal{Y}} r^{-1}(\{y\})$.
7. Calculation can be done summing over original space
   a. One need not first construct the distribution for the new variable.
   b. $\mathsf{E}(r(X)) = \sum_{x\in\mathcal{X}} r(x) p_X(x)$

i.
$$\sum_{y\in\mathcal{Y}} yp_Y(y) = \sum_{y\in\mathcal{Y}} y \sum_{x\in r^{-1}(\{y\})} p_X(x) = \sum_{y\in\mathcal{Y}} \sum_{x\in r^{-1}(\{y\})} r(x)p_X(x) = \sum_{x\in\mathcal{X}} r(x)p_X(x)$$

8. Linearity
   a. Let $Y = aX + b$ for some constants a, b
   b. Use transformation rule to show $\mathsf{E}(Y) = a\mathsf{E}(X) + b$.
      i. $\mathsf{E}(Y) = \sum_{x\in\mathcal{X}}(ax+b)p_X(x) = \sum_{x\in\mathcal{X}} axp_X(x) + \sum_{x\in\mathcal{X}} bp_X(x) = a\sum_{x\in\mathcal{X}} xp_X(x) + b\sum_{x\in\mathcal{X}} p_X(x) = a\mathsf{E}(X) + b$.
9. Other moments defined:
   a. The expectation is often referred to as the *first moment* of a random variable X;
   b. The r-*th moment* is defined as $\mathsf{E}(X^r)$.
   c. The r-*th central moment* is defined as $\mathsf{E}((X - \mathsf{E}(X))^r)$.
10. Describing spread
   a. *Variance*: $\mathsf{V}(X)$ is the second central moment: average squared distance from mean.
      i. $\mathsf{V}(X) = \mathsf{E}\left((X - \mathsf{E}(X))^2\right)$
      ii. Alternate formulation:
         - Square out what's inside: $\mathsf{V}(X) = \mathsf{E}\left(X^2 - 2X\mathsf{E}(X) + \mathsf{E}(X)^2\right)$

         - Break up using linearity: $\mathsf{V}(X) = \mathsf{E}\left(X^2\right) - 2\mathsf{E}(X\mathsf{E}(X)) + \mathsf{E}\left(\mathsf{E}(X)^2\right)$.
         - Pull out constants: $\mathsf{V}(X) = \mathsf{E}\left(X^2\right) - 2\mathsf{E}(X)^2 + \mathsf{E}(X)^2$
      iii. Obtain $\mathsf{V}(X) = \mathsf{E}\left(X^2\right) - \mathsf{E}(X)^2$.
   b. *standard deviation*: average distance from expectation: $\mathsf{SD}(X) = \sqrt{\mathsf{V}(X)}$
   c. Scaling: $\mathsf{V}(aX+b) = a^2\mathsf{V}(X)$.
      i.
$$\mathsf{V}(aX+b) = \mathsf{E}\left((aX + b - \mathsf{E}(aX+b))^2\right) = \mathsf{E}\left((aX + b - \mathsf{E}(aX) - b)^2\right) = \mathsf{E}\left(a^2(X - \mathsf{E}(X))^2\right) = a^2\mathsf{V}(X)$$
   d. Hence $\mathsf{SD}(aX+b) = |a|\,\mathsf{SD}(X)$
   e. Other dispersion measures: *mean absolute deviation* $\mathsf{E}(|X - \mathsf{E}(X)|)$ or $\mathsf{E}(|X - \text{median}(X)|)$.
      i. MAD scales the same way as SD, but will lack some useful properties later.

05

This page intentionally left blank.

This page intentionally left blank.