STAT 516: Discrete Random Variables

Lecture 5: CDF/pmf, Medians, Basic moments

Prof. Michael Levine

November 24, 2015

Discrete Random Variable Examples

- **Geometric** Random Variable
- **Binomial** Random Variable
- In general, each discrete random variable is described by its pmf

$$p_X(x) = P[X = x]$$

  for any $x \in \mathcal{D}$
- $p_X(x)$ always satisfies
  1. $0 \leq p_X(x) \leq 1$
  2. $\sum_{x \in \mathcal{D}} p_X(x) = 1$
- A **support** of a discrete random variable is a set of all points in $\mathcal{D}$ such that $p_X(x) > 0$

Hypergeometric Random Variable

- Each of N individuals can be characterized as a success (S) or failure (F), and there are M successes in the population
- A sample of n individuals is selected without replacement in such a way that each subset of size n is equally likely to be chosen.
- Let X be the number of Ss in a random sample of size n drawn from a population consisting of M Ss and $N - M$ Fs.
- The probability distribution of X, called the **hypergeometric distribution**, is given by

$$P(X = x) = h(x; n, M, N) = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}}$$

- In the above, $\max(0, n - N + M) \leq x \leq \min(n, M)$

Capture-recapture model

- 5 individuals from an animal population thought to be near extinction in a certain region have been caught, tagged and released.
- Afterwards, a sample of 10 animals is selected. Let X be the number of tagged animals in the second sample.
- The parameter values are $n = 10$, $M = 5$ and $N = 25$

Capture-Recapture Models

- Denote X the number of tagged animals in the *recapture* sample.
- The pmf of X is

$$h(x; 10, 5, 25) = \frac{\binom{5}{x}\binom{20}{10-x}}{\binom{25}{10}}$$

$$P(X = 2) = h(2; 10, 5, 25) = \frac{\binom{5}{2}\binom{20}{8}}{\binom{25}{10}} = .385$$

$$P(X \leq 2) = \sum_{x=0}^{2} h(x; 10, 5, 25) = .699$$

Computation

- To compute $P(X = 2)$ use $dhyper(2, m = 5, n = 20, k = 10)$
- To compute $P(X \leq 2)$, use $phyper(2, m = 5, n = 20, k = 10)$

One-to-one **Transformations**

- For X with $\mathcal{D}_X$ consider $Y = g(X)$
- Y has the range $\mathcal{D}_Y = \{g(x) : x \in \mathcal{D}_X\}$
- The pmf of Y is

$$p_Y(y) = p_X(g^{-1}(y))$$

Example

- For a geometric random variable X consider $Y = X - 1$
- If X is the flip number on which the first head appears, $p_X(x) = p(1-p)^{x-1}$
- Y is the number of failures before the first success
- $p_Y(y) = p_X(g^{-1}(y)) = p_X(y+1) = p(1-p)^x$

Example

- For $X \sim Bin\left(3, \frac{2}{3},\right)$ define $Y = X^2$
- $\mathcal{D}_X = \{x : x = 0, 1, 2, 3\}$ and $\mathcal{D}_Y = \{y : y = 0, 1, 4, 9\}$
- The inverse transformation is $g^{-1}(y) = \sqrt{y}$ which is a one-to-one in this case
- $p_Y(y) = p_X(\sqrt{y})$

Other transformations

- For a geometric random variable X, consider 1 unit gain when betting on odds and -1 when on evens
- The new variable is Y with $\mathcal{D}_Y = \{1, -1\}$
- Assume $\frac{1}{2}$;

$$p(X = 1, 3, 5, \ldots) = \sum_{x=1}^{\infty} \left(\frac{1}{2}\right)^{2x-1} = \frac{2}{3}$$

- Thus, $p_Y(1) = \frac{2}{3}$ and $p_Y(-1) = \frac{1}{3}$

Expectation

- We define $\mu = \mathbb{E}\, X = \sum_i x_i p(x_i)$ if $\sum_i |x_i| p(x_i) < \infty$
- If the sample space is finite or countably infinite,

$$\mu = \sum_x x p(x) = \sum_\omega X(\omega) P(\omega)$$

where $P(\omega)$ is the probability of a sample point ω

Expectation of a function of several random variables

- Let $X_1, \ldots, X_n$ be n discrete random variables on a common sample space Ω with a finite or a countably infinite number of sample points
- Assume that $\sum_{\omega} |g(X_1(\omega), X_2(\omega), \ldots, X_n(\omega)|P(\omega) < \infty$
- Define

$$\mathbb{E}\left[g(X_1, X_2, \ldots, X_n)\right) = \sum_{\omega} g(X_1(\omega), X_2(\omega), \ldots, X_n(\omega))P(\omega)$$

Example

- Consider two rolls of a fair die. Let X be the number of ones and Y the number of sixes obtained
- Define $g(X, Y) = XY$; note that $\Omega = \{11, 12, 13, \ldots, 64, 65, 66\}$ and $P(\omega) = \frac{1}{36}$
- $$\mathbb{E}(XY) = 0 \times \frac{1}{36} + \cdots = \frac{2}{36} = \frac{1}{18}$$

Basic properties of expectations

- For a finite constant c s.t. $P(X = c) = 1$ we have $\mathbb{E}\,X = c$
- For X and Y on the same sample space Ω with finite expectations, if $P(X \leq Y) = 1$, we have $\mathbb{E}\,X \leq \mathbb{E}\,Y$
- If X has a finite expectation, and $P(X \geq c) = 1$, $\mathbb{E}\,X \geq c$; if $P(X \leq c) = 1$, $\mathbb{E}\,X \leq c$

Linearity of expectations; expectation of a function

- Let $X_1, \ldots, X_n$ defined on the same Ω and $c_1, \ldots, c_k$ are any real valued constants. Then

$$\mathbb{E}\left(\sum_{i=1}^{k} c_i X_i\right) = \sum_{i=1}^{k} c_i \mathbb{E}\left(X_i\right)$$

- If X is defined on Ω and $Y = g(X)$, and if $\mathbb{E}\, Y$ exists,

$$\mathbb{E}\left(Y\right) = \sum_{\omega} Y(\omega) P(\omega) = \sum_{x} g(x) p(x)$$

Expectation of independent random variables

- Let $X_1, \ldots, X_k$ be independent random variables; if each expectation exists, we have

$$\mathbb{E}\,(X_1 X_2 \ldots X_k) = \mathbb{E}\,(X_1)\mathbb{E}\,(X_2)\cdots\mathbb{E}\,(X_k)$$

Example

- Let X be the sum of two rolls when a fair die is rolled twice
- Check that pmf of X is $p(2) = p(12) = \frac{1}{36}$; $p(3) = p(11) = \frac{2}{36}$ etc.
- $\mathbb{E}\,X = 2\frac{1}{36} + 3\frac{2}{36} + \cdots = 7$
- Alternatively, define X_1 the face on the first roll, X_2 the face on the second roll, then

$$\mathbb{E}\,X = \mathbb{E}\,(X_1 + X_2) = \mathbb{E}\,X_1 + \mathbb{E}\,X_2 = 3.5 + 3.5 = 7$$

Example

- Let a fair die be rolled 10 times and X be the sum of these rolls
- The pmf is hard to write down exactly; but if X_i be the face on ith roll,

$$\mathbb{E}\,X = \mathbb{E}\left(X_1 + X_2 + \cdots + X_{10}\right)$$
$$= \mathbb{E}\left(X_1\right) + \mathbb{E}\left(X_2\right) + \cdots + \mathbb{E}\left(X_{10}\right) = 3.5 \times 10 = 35$$

Use of indicator variables to compute expectations of discrete random variables

- Let $c_1, \ldots, c_m$ be constants and $A_1, \ldots, A_m$ some events
- Let X be an integer valued random variable $X = \sum_{i=1}^{m} c_i I_{A_i}$; then

$$\mathbb{E}\, X = \sum_{i=1}^{m} c_i P(A_i)$$

- Coin tosses - let a fair coin be tossed n times with the probability of success p
- The number of successes is $X = \sum_{i=1}^{n} I_{A_i}$ for obvious A_i
- $\mathbb{E}\, X = \sum_{i=1}^{m} P(A_i) = np$

Example

- The matching problem: let the number at location i be $p(i)$; define X as the number of locations such that $p(i) = i$
- Again, define obvious A_i and $X = \sum_{i=1}^{n} I_{A_i}$
- For any i, $P(A_i) = \frac{(n-1)!}{n!} = \frac{1}{n}$
- Thus, $\mathbb{E}\,X = \sum_{i=1}^{n} P(A_i) = 1$

Example

- Let a fair die be rolled n times; X is the number of faces that never show up in these n rolls
- Let A_i be the event that ith face is missing; $X = \sum_{i=1}^{6} I_{A_i}$
- For any i, $P(A_i) = \left(\frac{5}{6}\right)^n$
- $\mathbb{E}\, X = \sum_{i=1}^{6} P(A_i) = 6 \times \left(\frac{5}{6}\right)^n$
- If $n = 10$, this is about 0.97

Tail sum method

- Let X take values $0, 1, 2, \ldots$. Then,

$$\mathbb{E}\,X = \sum_{n=0}^{\infty} P(X > n)$$

- Let p be the probability of success in a Bernoulli trial; how long do we wait on average for the first success?
- If X is the number of trials needed than $X > n$ means that the first n trials all resulted in tails
- $\mathbb{E}\,X = \sum_{n=0}^{\infty} P(X > n) = \sum_{n=0}^{\infty} (1-p)^n = \frac{1}{p}$

Example

- Suppose a couple will have children until they have one child of each sex. How many children can they expect to have?
- Let X be the childbirth at which they have a child of each sex for the first time
- If the births are independent, the probability that a childbirth is a boy is p, we have

$$P(X > n) = p^n + (1-p)^n$$

- Therefore,

$$\mathbb{E}\,X = 2 + \sum_{n=2}^{\infty}[p^n + (1-p)^n] = 2 + \frac{p^2}{1-p} + \frac{(1-p)^2}{p} = \frac{1}{p(1-p)} - 1$$

Variance of discrete random variables

- **Variance** of X is
$$\sigma^2 = \mathbb{E}\left[(X-\mu)^2\right]$$
- The standard deviation is $\sigma = +\sqrt{\sigma^2}$
- Possible alternative is $\mathbb{E}\left|X-\mu\right|$...can show that $\sigma \geq \left|X-\mu\right|$

Basic properties

- $Var\left(cX\right) = c^2 var\left(X\right)$
- $Var\left(X+k\right) = Var\left(X\right)$ for any real k
- $Var\left(X\right) \geq 0$; $Var\left(X\right) = 0$ iff $X = \mu$ w.p.1
- $Var\left(X\right) = \mathbb{E}\left(X^2\right) - \mu^2$

Moments of a discrete random variable X

- For a positive integer $k \geq 1$ we call $\mathbb{E}\left(X^k\right)$ a **kth moment** of X
- $\mathbb{E}\left(X^{-k}\right)$ is a **kth inverse moment** of X
- If $\mathbb{E}\left[|X|^3\right] < \infty$, the **skewness** of X is

$$\beta = \frac{\mathbb{E}\left[(X-\mu)^3\right]}{\sigma^3}$$

- The skewness measures how symmetric the distribution if X is; e.g. for $X \sim N(0,1)$, $\beta = 0$
- If $\mathbb{E}\left[X^4\right] < \infty$, the **kurtosis** of X is

$$\gamma = \frac{\mathbb{E}\left[(X-\mu)^4\right]}{\sigma^4} - 3$$

- The kurtosis is always ≥ -2 and is equal to zero for a normal distribution; it measures how "spiky" the distribution is around its mean

Example

- Let X be the sum of two independent rolls of a fair die. We know that $\mathbb{E}(X) = 7$
- $\mathbb{E}(X^2) = \frac{329}{6}$ and
  $Var(X) = \mathbb{E}(X^2) - \mu^2 = \frac{329}{6} - 49 = \frac{35}{6} = 5.83$

Example: Variance in the Matching Problem

- Let X be the number of locations where match occurs when n numbers are rearranged in a random order
- We know that $\mathbb{E}(X) = 1$ for any n; moreover, recall representation $X = \sum_{i=1}^{n} I_{A_i}$
- Now,

$$\mathbb{E}(X^2) = \sum_{i=1}^{n} P(A_i) + 2 \sum_{1 \le i < j \le n} P(A_i \cap A_j)$$
$$= n \times \frac{1}{n} + 2 \binom{n}{2} \frac{(n-2)!}{n!} = 1 + 1 = 2$$

- Thus, $Var(X) = 2 - 1 = 1$ for any n
- Think of which distribution might approximate well the number of matches...

Variance of a sum of independent random variables

- Let $X_1, \ldots, X_n$ be independent random variables;

$$Var\left(\sum_{i=1}^{n} X_i\right) = \sum_{i=1}^{n} Var\left(X_i\right)$$

- As a corollary (and a very important one!), if $\bar{X} = \frac{\sum_{i=1}^{n} X_i}{n}$, and $\sigma^2 = Var\, X_i < \infty$,

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

Chebyshev's and Markov's inequalities

- (Chebyshev's inequality) Let $\mathbb{E}\, X = \mu$ and $Var\, X = \sigma^2$ be finite. For any positive number k we have

$$P(|X - \mu| \geq k\sigma) \leq \frac{1}{k^2}$$

- (Markov's inequality) Suppose X takes only nonnegative values and $\mathbb{E}\, X = \mu$ is finite. If c is any positive number,

$$P(X \geq c) \leq \frac{\mu}{c}$$

- Chebyshev's inequality is a direct consequence of Markov's inequality

Example

- Chebyshev's and Markov's inequalities are rather conservative
- X be the sum of two rolls of a fair die. Recall that $\mu = 7$ and $\sigma = 2.415$
- Choose $k = 2$ in the Chebyshev's inequality; direct calculation gives

$$P(|X - 7| \geq 4.830) = \frac{1}{18} = 0.056$$

- Chebyshev's inequality gives the lower bound of $\frac{1}{4} = 0.25$

Weak law of large numbers

- WLLN is a direct consequence of Chebyshev's inequality
- Let $X_1, \ldots, X_n$ be iid RV's with $\mathbb{E}\, X_i = \mu$ and $Var\, X_i = \sigma^2 < \infty$.
- For any $\varepsilon > 0$,
$$P(|\bar{X} - \mu| > \varepsilon) \to 0$$
as $n \to \infty$
- A stronger version is the strong law of large numbers (SLLN) that says that
$$P(\lim_{n\to\infty} \bar{X} = \mu) = 1$$
- The only conditions needed is that $\mathbb{E}\,|X_i|$ be finite

Truncated Distributions

- Examples: planet observations; reported car accidents
- Let X be a discrete random variable with pmf $p(x)$; let A be a fixed subset of its values
- **The distribution of** X **truncated to** A has the pmf

$$p_A(y) = \frac{p(y)}{P(X \in A)}$$

  for any $y \in A$ and 0 if $y \notin A$
- The mean of a truncated distribution is

$$\mu_A = \frac{\sum_{y \in A} y p(y)}{\sum_{y \in A} p(y)}$$

Example

- Let $P(X = n) = \frac{1}{2^n}$ for $n = 1, 2, \ldots$; we only observe X if $X \leq 5$
- The truncation set is $A = \{1, 2, 3, 4, 5\}$ and

$$p_A(y) = \frac{(1/2)^y}{\sum_{y=1}^{5}(1/2)^y} = \frac{2^{5-y}}{31}$$

  for $y = 1, 2, \ldots, 5$
- Check that its mean is 1.71 which is less than $\sum_{n=1}^{\infty} n \times \frac{1}{2^n} = 2$
- (Chow-Studden inequality). For any RV X and finite real constants a and b, let $U = \min(X, a)$ and $V = \max(X, b)$. Then,

$$Var\,(U) \leq Var\,(X);\, Var\,(V) \leq Var\,(X)$$