Chapter 2: Transformations and Expectations

MATH 667-01
Statistical Inference
University of Louisville
Textbook: Statistical Inference by Casella and Berger

August 27, 2015
Last corrected: 9/6/15

- Suppose X is a random variable with sample space $\mathcal{X}$ and associate Borel field $\mathcal{B}$.
- Suppose $g(x)$ is a function with domain $\mathcal{X}$. Then $Y = g(X)$ is a new random variable with sample space $\mathcal{Y}$.
- Consider the inverse mapping $g^{-1}(A) = \{x \in \mathcal{X} : g(x) \in A\}$.
- For sets A such that $g^{-1}(A) \in \mathcal{B}$, probabilities for Y can be determined using

$$P(Y \in A) = P(g(X) \in A) = P(X \in g^{-1}(A)).$$

- If X is a discrete random variable with pmf $f_X(x)$, then $Y = g(X)$ is also a discrete random variable.
- The sample space of Y is $\mathcal{Y} = \{y : y = g(x) \text{ for some } x \in \mathcal{X}\}$.
- For $y \in \mathcal{Y}$, the pmf of Y is given by

$$\begin{aligned} f_Y(y) &= P(Y = y) \\ &= \sum_{x \in g^{-1}(y)} P(X = x) \\ &= \sum_{x \in g^{-1}(y)} f_X(x). \end{aligned}$$

- *Example*: Suppose X has pmf

$$f_X(x) = \frac{1}{6}\left(\frac{5}{6}\right)^{x-1} \quad \text{for } x \in \mathcal{X}$$

where $\mathcal{X}$ is the set of positive integers and let $Y = (-1)^X$. Then $\mathcal{Y} = \{-1, 1\}$ and the pmf of Y at -1 is

$$\begin{aligned} f_Y(-1) = \sum_{x \text{ is odd}} f_X(x) &= \frac{1}{6} + \frac{1}{6}\left(\frac{5}{6}\right)^2 + \frac{1}{6}\left(\frac{5}{6}\right)^4 + \dots \\ &= \frac{1}{6}\left(\frac{1}{1-(5/6)^2}\right) = \frac{6}{11}. \end{aligned}$$

- Suppose that X is a random variable with pdf/pmf $f_X(x)$ and $\mathcal{X} = \{x : f_X(x) > 0\}$.
- Let $\mathcal{Y} = \{y : y = g(x) \text{ for some } x \in \mathcal{X}\}$.
- *Theorem*: Let X have cdf $F_X(x)$, let $Y = g(X)$.
  a. If g is an increasing function on $\mathcal{X}$, $F_Y(y) = F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.
  b. If g is a decreasing function on $\mathcal{X}$ and X is a continuous random variable, $F_Y(y) = 1 - F_X(g^{-1}(y))$ for $y \in \mathcal{Y}$.
- *Theorem*: Let X have pdf $f_X(x)$ and let $Y = g(X)$, where g is a monotone function. Suppose that $f_X(x)$ is continuous on $\mathcal{X}$ and that $g^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$. Then the pdf of Y is given by

$$f_Y(y) = \begin{cases} f_X(g^{-1}(y)) \left| \frac{d}{dy} g^{-1}(y) \right| & y \in \mathcal{Y} \\ 0 & \text{otherwise} \end{cases} .$$

- *Theorem*: Let X have pdf $f_X(x)$, let $Y = g(X)$ and define $\mathcal{X} = \{x : f_X(x) > 0\}$. Suppose there exists a partition, $A_0, A_1, \ldots, A_k$, of $\mathcal{X}$ such that $P(X \in A_0) = 0$ and $f_X(x)$ is continuous on each A_i. Further, suppose there exist functions $g_1(x), \ldots, g_k(x)$ defined on $A_1, \ldots, A_k$, respectively, satisfying
  a. $g(x) = g_i(x)$, for $x \in A_i$,
  b. $g_i(x)$ is monotone on A_i,
  c. the set $\mathcal{Y} = \{y : y = g_i(x) \text{ for some } x \in \mathcal{X}\}$ is the same for each $i = 1, \ldots, k$, and
  d. $g_i^{-1}(y)$ has a continuous derivative on $\mathcal{Y}$, for each $i = 1, \ldots, k$.

  Then

$$f_Y(y) = \begin{cases} \sum_{i=1}^{k} f_X(g_i^{-1}(y)) \left| \frac{d}{dy} g_i^{-1}(y) \right| & y \in \mathcal{Y} \\ 0 & \text{otherwise} \end{cases}.$$

- *Example*: Let X be a standard normal random variable with pdf
$$f_X(x) = \frac{1}{\sqrt{2\pi}} e^{-x^2/2}, -\infty < x < \infty.$$
Find the pdf of $Y = X^2$.
- *Answer*: The pdf of Y is
$$f_Y(y) = \frac{1}{\sqrt{2\pi}} y^{-1/2} e^{-y/2}, 0 < y < \infty.$$
(This is a chi squared random variable with 1 degree of freedom.)

- *Definition*: The *expected value* or *mean* of a random variable $g(X)$, denoted by $\mathrm{E}g(X)$, is

$$\mathrm{E}g(X) = \begin{cases} \int_{-\infty}^{\infty} g(x) f_X(x)\ dx & \text{if } X \text{ is continuous} \\ \sum_{x \in \mathcal{X}} g(x) f_X(x) & \text{if } X \text{ is discrete} \end{cases}$$

  provided that the integral or the sum exists.
  If $\mathrm{E}|g(X)| = \infty$, we say that $\mathrm{E}g(X)$ does not exist.
- *Example*: Show that $\mathrm{E}X$ does not exist if X is a *Cauchy random variable* with pdf

$$f_X(x) = \frac{1}{\pi(1 + x^2)}, -\infty < x < \infty.$$

- *Theorem*: Let X be a random variable and let a, b, and c be constants. Then for any functions $g_1(x)$ and $g_2(x)$ whose expectations exist,
  a. $\mathsf{E}(ag_1(X) + bg_2(X) + c) = a\mathsf{E}g_1(X) + b\mathsf{E}g_2(X) + c$.
  b. If $g_1(x) \geq 0$ for all x, then $\mathsf{E}g_1(X) \geq 0$.
  c. If $g_1(x) \geq g_2(x)$ for all x, then $\mathsf{E}g_1(X) \geq \mathsf{E}g_2(X)$.
  d. If $a \leq g_1(x) \leq b$ for all x, then $a \leq \mathsf{E}g_1(X) \leq b$.
- *Example*: Find the real number b which minimizes the expected value of $(X - b)^2$.

- *Definition*: For each integer n, the *nth moment of X* (or $F_X(x)$), is $\mu'_n = \mathsf{E}X^n$.
  The *nth central moment of X* is $\mu_n = \mathsf{E}(X-\mu)^n$, where $\mu = \mu'_1 = \mathsf{E}X$ is referred to as the *mean*.
- *Definition*: The *variance* of a random variable X is its second central moment, $\text{Var } X = \mathsf{E}(X - \mathsf{E}X)^2$.
  The *standard deviation* of X is $\sqrt{\text{Var } X}$.
- An useful alternative formula for the variance is

$$\text{Var } X = \mathsf{E}X^2 - (\mathsf{E}X)^2.$$

- *Theorem*: If X is a random variable with finite variance, then for any constants a and b, $\text{Var}(aX+b) = a^2 \text{Var } X$.

- *Example*: (a) Show that the mean and variance of a Poisson random variable X with pmf

$$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, \ldots$$

  is λ.

  (b) If X is a Poisson random variable with mean 7, find the mean and variance of $2X - 11$.

- *Definition*: Let X be a random variable with cdf F_X. The *moment generating function (mgf) of X* (or F_X) is $M_X(t) = \mathsf{E}e^{tX}$, provided that the expectation exists for t in some neighborhood of 0 (that is, there is an $h > 0$ such that $\mathsf{E}e^{tX}$ exists for all t in $-h < t < h$). If the expectation does not exist in a neighborhood of 0, we say that the moment generating function does not exist.

- *Theorem*: For any constants a and b, the mgf of the random variable $aX + b$ is given by

$$M_{aX+b}(t) = e^{bt} M_X(at).$$

- *Theorem*: If X has mgf $M_X(t)$, then $\mathsf{E}X^n = M_X^{(n)}(0)$, where we define $M_X^{(n)}(0) = \dfrac{d^n}{dt^n} M_X(t)|_{t=0}$.

- *Example*: (a) Show that the mgf of a Poisson random variable X with pmf

$$P(X = x|\lambda) = \frac{e^{-\lambda}\lambda^x}{x!}, x = 0, 1, \ldots$$

is $M_X(t) = \exp\left\{\lambda(e^t - 1)\right\}$.

(b) If X is a Poisson random variable with mean 7, find the mgf of $2X - 11$.

- *Theorem*: Let $F_X(x)$ and $F_Y(y)$ be two cdfs all of whose moments exist.
  a. If F_X and F_Y have bounded support, then $F_X(u) = F_Y(u)$ for all u if and only if $\mathsf{E}X^r = \mathsf{E}Y^r$ for all integers $r = 0, 1, 2, \ldots$.
  b. If the moment generating functions exist and $M_X(t) = M_Y(t)$ for all t in some neighborhood of 0, then $F_X(u) = F_Y(u)$ for all u.

- *Theorem*(Convergence of mgfs): Suppose $\{X_i, i = 1, 2, \ldots\}$ is a sequence of random variables, each with mgf $M_{X_i}(t)$. Furthermore, suppose that $\lim_{i \to \infty} M_{X_i}(t) = M_X(t)$, for all t in a neighborhood of 0, and $M_X(t)$ is an mgf. Then there is a unique cdf F_X whose moments are determined by $M_X(t)$ and, for all X where $F_X(x)$ is continuous, we have $\lim_{i \to \infty} F_{X_i}(x) = F_X(x)$.

- *Example*: The mgf of a binomial(n,p) random variable X_n is

$$M_{X_n}(t) = \left((1-p) + pe^t\right)^n .$$

If $p = \lambda/n$, then

$$\begin{aligned} M_{X_n}(t) &= \left(\left(1 - \frac{\lambda}{n}\right) + \frac{\lambda}{n}e^t\right)^n \\ &= \left(1 + \frac{\lambda}{n}(e^t - 1)\right)^n \to \exp\left\{\lambda(e^t - 1)\right\} \end{aligned}$$

as $n \to \infty$. Since $M_X(t) = \exp\left\{\lambda(e^t - 1)\right\}$ is the mgf of a Poisson(λ) random variable, the sequence of distribution functions of X_n converges to the distribution function of a Poisson random variable at all of its continuity points.

- The *characteristic function of* a random variable X is $\varphi_X(t) = \mathsf{E}e^{itX}$, where $i = \sqrt{-1}$.
- *Theorem*(Convergence of characteristic functions): Suppose $\{X_i, i = 1, 2, \dots\}$ is a sequence of random variables, each with characteristic function $\varphi_{X_i}(t)$. Furthermore, suppose that $\lim_{i \to \infty} \varphi_{X_i}(t) = \varphi_X(t)$, for all t in a neighborhood of 0, and $\varphi_X(t)$ is a characteristic function. Then there is a unique cdf F_X whose moments are determined by $\varphi_X(t)$ and, for all X where $F_X(x)$ is continuous, we have $\lim_{i \to \infty} F_{X_i}(x) = F_X(x)$.