Moment generating functions

More applications (cont'd)

3. Convergence (more on this later)

- Suppose a r.v. X has mgf $M_X(t)$ and suppose $X_1, X_2, \ldots$ are a sequence of random variables, where X_n has mgf $M_{X_n}(t)$ for each $n \geq 1$.
  If
  $$\lim_{n\to\infty} M_{X_n}(t) = M_X(t)$$
  holds for all $t \in (-h, h)$ for some $h > 0$, then the sequence $X_1, X_2, \ldots$ converges (in distribution) to X.

$$\lim_{x\to\infty}(f(x))^{g(x)} = 1^{\infty} \Longrightarrow \lim_{x\to\infty}(f(x))^{g(x)} = e^{\lim_{x\to\infty}(f(x)-1)g(x)}$$

Example: Suppose X_n pmf $f_X(x) = \binom{n}{x} p^x (1-p)^{n-x}$, $x = 0, 1, \ldots, n$.

Then X_n is a Binomial(n, p) r.v., counting the number of "successes" in n separate trials (more later).

Suppose that $n \to \infty$ and $\lambda = np$ remains constant.

Recall: If $X \sim Bin(n,p) \Rightarrow M_X(t) = (pe^t + 1 - p)^n$

big $n \uparrow$ $\quad p_{\text{small}} = \frac{\lambda}{n}$, $p \to 0$

$$\lim_{n\to\infty} M_{X_n}(t) = \lim_{n\to\infty}(pe^t + 1 - p)^n = \lim_{n\to\infty}(1 + p(e^t - 1))^n$$
$$= \lim_{n\to\infty}\left(1 + \frac{\lambda(e^t - 1)}{n}\right)^n$$
$$= e^{\lambda(e^t - 1)}$$

Recall: $\boxed{\lim_{n\to\infty}\left(1 + \frac{a}{n}\right)^n = e^a}$ (*)

Note: Poisson distribution is often used to model counts of rare events. A random variable X having Poisson distribution has $\mathbb{P}(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$ $\quad x = 0, 1, 2, \ldots$

pmf $\quad f(x) = e^{-\lambda}\frac{\lambda^x}{x!}$, $x = 0, 1, 2, 3, \ldots$ $\quad$ & $\quad$ mgf $\quad M_X(t) = \mathrm{E}e^{tX} = e^{\lambda(e^{\lambda}-1)}$, $t \in \mathbb{R}$

$$\lim_{n\to\infty} M_{X_n}(t) = e^{\lambda(e^t - 1)} \longrightarrow \text{MGF of a Poisson}(\lambda)$$

(ASIDE) $\to M_X(t) = \mathbb{E}[e^{tX}] \to$ MGF

Moment generating functions

Other generating functions

There is nothing particularly illuminating about mgf's: these are just a technical device that are sometimes useful for proving theorems. In fact, they are not the only transforms for technical purposes or even the most ~~useful~~ (e.g., Fourier transforms/chacteristic functions are often more useful in STAT 642).

$$\left.\frac{d^n M_X(t)}{dt^n}\right|_{t=0} = \mathbb{E}(X^n)$$

1. $\log M_X(t)$ is the cumulant generating function

$$\left.\frac{d^n \log M_X(t)}{dt^n}\right|_{t=0} = n\text{th cumulant}$$

where 1st cumulant is $\mathrm{E}X$ & 2nd cumulant is $\mathrm{Var}(X)$

$M_X(t) = \mathbb{E}[e^{tX}]$

2. $\mathrm{E}t^X$ is the factorial moment generating function (fmgf)

$$\boxed{\left.\frac{d^n \mathrm{E}t^X}{dt^n}\right|_{t=1} = \mathrm{E}X(X-1)\cdots(X-n+1)}$$

For a discrete r.v. the fmgf is also called the probability generating function

$n=2 \Rightarrow d^2 E$

3. $\phi_X(t) = \mathrm{E}e^{\imath tX} = \mathrm{E}\cos(tX) + \imath\mathrm{E}\sin(tX)$ is the characteristic function
   - above $e^{\imath tx} = \cos(tx) + \imath\sin(tx)$, $\imath = \sqrt{-1}$ (i.e., requires complex numbers)
   - always exists for all $t \in \mathbb{R}$
   - uniquely determines the distribution of X
   - derivatives can be used to obtain moments

$e^{i\theta} = \cos\theta + i\sin\theta$

Moment generating functions

One last thing

We have several times in lecture up to this point interchanged “orders of limits” (i.e., switch the order of derivatives and expectation or derivatives and summations or derivatives and integrals). This interchange is also implicit in using mgf’s to compute moments:

$$
\begin{aligned}
\left.\frac{d^n M_X(t)}{dt^n}\right|_{t=0} &\overset{1}{=} \left.\frac{d^n \mathrm{E} e^{tX}}{dt^n}\right|_{t=0} \\
&\overset{2}{=} \mathrm{E}\left(\left.\frac{d^n e^{tX}}{dt^n}\right|_{t=0} \right) \\
&= \mathrm{E}\left(\left. X^n e^{tX}\right|_{t=0} \right) \\
&= \mathrm{E} X^n
\end{aligned}
$$

We don’t need to worry about the validity of these interchanges here (it’s covered in STAT 642 where it makes more sense with the right technical material in hand).

Section 2.4 of Casella & Berger addresses the issue of conditions under which such interchanges are permissible; this is for mathematical completeness.

Inequalities

Markov inequality

- So far, we have discussed computing exact probabilities
- Inequalities can provide insight with less calculation and have theoretical importance/uses

  This material can also be found in Section 3.6, Casella & Berger (though this has nothing to do with two or more random variables at a time)

- **Markov inequality:** Suppose X is a r.v. and $\boxed{g(x) \geq 0.}$ Then, for any $r > 0$,

$$\mathbb{P}(Y \in A) = \int_A f_Y(y)\,dy$$

$$\boxed{P(g(X) \geq r) \leq \frac{\mathrm{E}g(X)}{r}.}$$

$$g(x) \geq r \qquad \frac{g(x)}{r} \geq 1$$

Proof.

$$\mathbb{P}(g(X) \geq r) = \int_{\{x:\, g(x) \geq r\}} 1 \cdot f_X(x)\,dx \leq \int_{\{x:\, g(x) \geq r\}} \frac{g(x)}{r} f_X(x)\,dx$$

$$\leq \int_{\mathbb{R}} \frac{g(x)}{r} f_X(x)\,dx$$

$$= \frac{1}{r}\int_{\mathbb{R}} g(x) f_X(x)\,dx = \frac{1}{r}\mathbb{E}[g(X)].$$

- Examples
  - If $g(x) = |x|$ and $r > 0$, then

$$P(|X| \geq r) \leq \frac{\mathrm{E}|X|}{r}$$

  - If $g(x) = |x|^n$ ($n \geq 1$) and $r > 0$, then

$$P(|X| \geq r) = P(|X|^n \geq r^n) \leq \frac{\mathrm{E}|X|^n}{r^n} \quad (*)$$

Note: X is a r.v. with $\mathbb{E}(X) = 0$

$n = 2$, $\mathbb{P}(|X| \geq r) \leq \frac{\mathbb{E}X^2}{r^2} = \frac{Var(X)}{r^2}$

  - If $g(x) = e^{tx} > 0$ for $t > 0$ and $a \in \mathbb{R}$,

$$P(X \geq a) = P(e^{tX} \geq e^{at}) \leq \frac{\mathrm{E}e^{tX}}{e^{at}} = \frac{M_X(t)}{e^{at}}$$

ex: $\mathbb{P}(|X-\mu|>a) \geq \cdots$

Inequalities

$\mathbb{P}(|X-\mu|$

Chebychev inequality

- Bounds the probability that a r.v. X falls (or doesn't) in a neighborhood $\mu \pm \sigma k$ around its mean $\mathrm{E}X = \mu$, where $\sigma = \sqrt{\mathrm{Var}(X)}$ is the standard deviation

- Special case of the Markov inequality

- **Chebychev inequality:** Suppose X is a r.v. with mean $\mathrm{E}X = \mu$ and variance σ^2. Then, for any $k > 0$,

$\mathbb{P}(A^C) = 1 - \mathbb{P}(A)$

$$P(\overbrace{|X-\mu| \geq k\sigma}^{A}) \leq \frac{1}{k^2} \qquad \& \qquad P(\overbrace{|X-\mu| < k\sigma}^{A^C}) \geq 1 - \frac{1}{k^2}$$

Proof. Set $g(x) = (x-\mu)^2/\sigma^2$ and note $\mathrm{E}g(X) = 1$. Then,

$$\mathbb{E}[g(X)] = \mathbb{E}\left[\frac{(X-\mu)^2}{\sigma^2}\right] = \frac{1}{\sigma^2}\,\mathbb{E}[(X-\mu)^2] \overset{\text{def}}{=} \frac{\sigma^2 = \mathrm{Var}(X)}{\sigma^2} = 1$$

$$\begin{aligned} P(|X-\mu| \geq k\sigma) = P\left(\frac{|X-\mu|}{\sigma} \geq k\right) &= P\left(\frac{|X-\mu|^2}{\sigma^2} \geq k^2\right) \\ &= P(g(X) \geq k^2) \\ &\leq \frac{\mathrm{E}g(X)}{k^2} = \frac{1}{k^2} \end{aligned}$$

- Example
  - If $k = 2$, $\mathbb{P}(|X-\mu| \geq k\sigma) \leq \frac{1}{k^2}$

    $$P(|X-\mu| \geq \underset{k=2}{2}\sigma) \leq \underset{\frac{1}{2^2}}{0.25} \qquad P(|X-\mu| < 2\sigma) \geq 0.75$$

  - If $k = 3$,

    $$P(|X-\mu| \geq 3\sigma) \leq \underbrace{0.11}_{\frac{1}{3^2}} \qquad P(|X-\mu| < 3\sigma) \geq 0.89$$

  - If $k = 4$,

    $$P(|X-\mu| \geq 4\sigma) \leq \underbrace{0.0625}_{\frac{1}{4^2}} \qquad P(|X-\mu| < 4\sigma) \geq 0.9375$$

Inequalities

Jensen's inequality

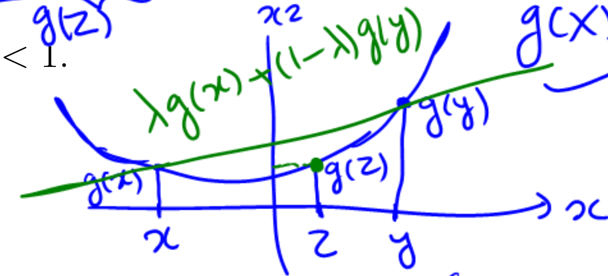
- Definition of a convex function: $g(x)$ is **convex** if

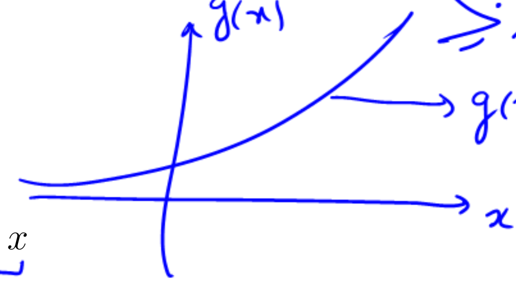
$$g(\underbrace{\lambda x + (1-\lambda) y}_{z}) \le \lambda g(x) + (1-\lambda) g(y)$$

for all x, y and all $0 < \lambda < 1$.

$g(X) = X^2$

Verify $g(\lambda x + (1-\lambda)y) \stackrel{\text{def}}{=} (\lambda x + (1-\lambda)y)^2 \stackrel{?}{\ge} \lambda x^2 + (1-\lambda) y^2$





- Also, $g(x)$ is convex if $g''(x) \ge 0$ for all x

$g(x) = x^2 \rightarrow g'(x) = 2x \Rightarrow g''(x) = 2 > 0$

- $g(x)$ is **concave** if $-g(x)$ is convex

$g(x) = -\log x \Rightarrow g'(x) = \frac{-1}{x} \Rightarrow g''(x) = \frac{1}{x^2} > 0$

- **Jensen's inequality:** Suppose X is a r.v. and $g(x)$ is a convex function. Then,

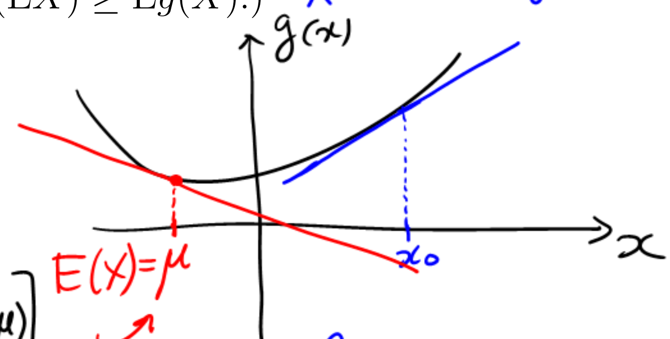
$$\boxed{\mathrm{E} g(X) \ge g(\mathrm{E}X)}$$

$\mathrm{E}X = \mu$

(For a concave function, the reverse holds: $g(\mathrm{E}X) \ge \mathrm{E}g(X)$.)

Proof: $\mathbb{E}(g(X)) \stackrel{?}{\ge} g(\mathbb{E}(X))$

$\rightarrow \ell(x) = g(\mu) + m(x - \mu)$, $m = g'(\mu)$



$g(x) \ge \ell(x)$

$\Rightarrow \mathbb{E}(g(X)) \ge \mathbb{E}(\ell(X)) = \mathbb{E}[g(\mu) + m(X - \mu)]$

$= \underbrace{\mathbb{E}[g(\mu)]}_{g(\mu)} + m \underbrace{\mathbb{E}[\cancel{(X-\mu)}]}_{0}$

$= g(E(X))$

$\ell(x) = a + bx$ is called the tangent function of g at point x_0

$\mathbb{E}[g(X)] \ge g(\mathbb{E}(X))$

Inequalities

Jensen's inequality: examples

1. $g(x) = x^2$ is convex so that

$$\mathrm{E}X^2 \geq (\mathrm{E}X)^2 \Rightarrow \mathrm{Var}(X) = \mathbb{E}(X^2) - (\mathbb{E}(X))^2 \geqslant 0$$

$\mathbb{E}[g(X)] = \mathbb{E}[X^2]$

$g(\mathbb{E}(X)) = (\mathbb{E}(X))^2$

$\mathbb{E}(g(X)) \leq g(\mathbb{E}(X))$

2. $g(x) = \log x$ is concave so that, assuming $X \geq 0$,

$\mathbb{E}(\log X) \leq \log(\mathbb{E}(X))$

$(g(x) = -\log x$ Convex$)$

$$\log(\mathrm{E}X) \geq \mathrm{E}\log X$$

✓

✓ 3. Can use Jensen's inequality to show that, for a set of real numbers $a_1, \ldots, a_n$,

$$\frac{1}{n}\sum_{i=1}^{n} |a_i| \geq \left(\prod_{i=1}^{n} |a_i|\right)^{1/n}$$

Hint: You can use (2) above $(g(X) = \log X)$

Define $\mathbb{P}(X = x) = \begin{cases} \frac{1}{n} & x = |a_1|, |a_2|, \ldots, |a_n| \\ 0 & \text{o.w.} \end{cases}$

$\mathbb{E}[X] = \sum x\,\mathbb{P}(X=x) = \dfrac{\sum_{i=1}^{n} |a_i|}{n}$

Compute $\log(\mathbb{E}(X))$

$\mathbb{E}(\log X)$