7 Moment Generating Functions

7.1 Moment Generating Functions

Slide 1

Moment Generating Functions

We introduce a useful tool for later arguments.

Moment Generating Function (mgf): The mgf for a r.v. X is

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

The mgf encaptures all moments of X:

Theorem 4.1: Let X be a r.v. with mgf $M_X(t)$. If exchanging expectation and differentiation is valid, then

$$E(X^n) = M_X^{(n)}(0)$$

where M(n)(0) is the nth derivative of MX (t) evaluated at t = 0.

Proof: under the assumption, we get

$$M_X^{(n)}(t) = E[X^n e^{tX}] \implies M_X^{(n)}(0) = E(X^n).$$

Slide 2

Moment Generating Function (ctd.)

Pr(X = n) = (1 – p)^(n–1)*p

Example: $X \sim \text{Geom}(p)$. Then, letting $q = 1 - p$, for $qe^t < 1$, or $t < -\log(q)$[1]

$$M_X(t) = Ee^{tX} = \sum_{k=1}^{\infty} q^{k-1} pe^{tk} = \frac{p}{q}\sum_{k=1}^{\infty}(qe^t)^k = \frac{p}{q}\left(\frac{1}{1-qe^t} - 1\right).$$

Geometric series:
a/(1-r) = a + ar + ar^2 + ar^3 + …
In this case, a = 1, r = q*e^t

Theorem 4.2: If $M_X(t) = M_Y(t)$ for $-\delta < t < \delta$ for some $\delta > 0$, then X and Y have the same distribution.
(*without proof*)

Theorem 4.3: If $X \perp Y$ then $M_{X+Y}(t) = M_X(t)M_Y(t)$

Solution: exercise.

7.2 Chernoff Bounds

Slide 3 Markov's inequality: Pr(X ≥ a) ≤ E(X)/a

Chernoff Bounds

Use Markov's inequality for e^{tX}, substituting $E(e^{tX}) = M_X(t)$

$$\Pr(X \geq a) = \Pr(e^{tX} \geq e^{ta}) \leq \frac{E(e^{tX})}{e^{ta}}.$$

Right tail: This is true for all $t > 0$ – use the one with the sharpest bound:

$$\Pr(X \geq a) \leq \min_{t>0} \frac{M_X(t)}{e^{ta}}.$$

[1] the constraint on t was missing in the lecture

Left tail: For $\Pr(X \leq a)$ use $t < 0$ and then again Markov's inequ:

$$\Pr(X \leq a) = \Pr(e^{tX} \geq e^{ta}) \leq \min_{t<0} \frac{M_X(t)}{e^{ta}}$$

Clever use of these allows many useful bounds that are better than Chebyshev bounds.

7.3 Examples

Example: Flipping a Coin
X ~ Bern(0.5)
P(X = 1) = 0.5 (probability of success, getting "Heads")
P(X = 0) = 0.5 (probability of failure, getting "Tails")

Slide 4

Examples

1. Find $M_X(t)$ for $X \sim \text{Bern}(p)$, i.e., $p_X(1) = p$.

P(X = 0) P(X = 1)

Solution: $M_X(t) = (1-p)\cdot 1 + p \cdot e^t = 1 + p(e^t - 1)$.

Using $1 + x \leq e^x$ we get

Taylor Series: e^x = 1 + x + x^2/2! +x^3/3! …

$$M_X(t) \leq e^{p(e^t-1)},$$

M (t) = 1+p(e^t–1)
= 1+p*e^t - p
= 1+p*e^t - (1-q)
= q + pe^t.

which is sometimes a useful bound.

Writing $q = 1 - p$ we get alternatively $M_X(t) = q + pe^t$.

2. Find $M_Y(t)$ for $Y \sim \text{Bin}(n, p)$. X ~ Bin(n, p) = (n j)p^j*(1-p)^(n-j)

Solution: Use $Y = \sum_{i=1}^n X_i$, with $X_i \sim \text{Bern}(p)$, and therefore, letting $q = 1 - p$, we get $M_Y(t) = (q + pe^t)^n$.

Slide 5

3. If $X \sim \text{Poi}(\lambda)$, show that for $i < \lambda$,

$$\Pr(X \leq i) \leq \frac{e^{-\lambda}(e\lambda)^i}{i^i}.$$

Hint: If $X \sim \text{Poi}(\lambda)$, then $M_X(t) = e^{\lambda(e^t-1)}$.

Solution: First,

$$\Pr(X \leq a) \leq M(t)/e^{ta} = e^{\lambda(e^t-1)-ta},$$

for $t < 0$.

To find the lowest bound we minimize $\lambda(e^t - 1) - ta$, to find $e^t = a/\lambda \implies t = \log(a/\lambda) < 0$ (for $a < \lambda$).

Substituting in the Chernoff bound we get

$$\Pr(X \leq a) \leq e^{\lambda(a/\lambda-1)}\left(\frac{\lambda}{a}\right)^a = \frac{e^{-\lambda}(e\lambda)^a}{a^a}$$

as claimed.

Slide 6

For those following along in the book: We will

- skip the rest bof §4, including §4.2.1 and beyond,

- skip §5-7, and
- continue with §8 in the next unit.

Some material (§4.3, §6.1 and &7.1) will be briefly introduced in homeworks.