Statistical Inference

Lecture 02a

ANU - RSFAS

Last Updated: Tue Feb 28 15:21:58 2017

Moment Generating Functions

Definition: (Rice Sec. 4.5) Let X be a random variable. The moment generating function (mgf) of X, denoted by $M_x(t)$ or just $M(t)$ is

$$M_x(t) = E\left(e^{tx}\right).$$

- Provided the expectation exists for t in a neighborhood of 0.
- Facts:
  1. **Property C**: $M_{a+bX}(t) = E\left(e^{t(a+bX)}\right) = e^{ta}E\left(e^{(bt)X}\right) = e^{ta}M_X(bt)$.
  2. **Property D**: $M_{X+Y}(t) = M_X(t)M_Y(t)$ if X and Y are independent. Why?

- Why is this called the MGF?
  - Note (math fact): $e^{tX} = 1 + \frac{tX}{1!} + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots$
  - $M(t) = 1 + tE(X) + \frac{t^2}{2!}E(X^2) + \frac{t^3}{3!}E(X^3) \cdots$
  - Let's differentiate $M(t)$ with respect to t . . . k times and set $t = 0$:

$$\frac{d^k}{dt^k} M(t)\bigg|_{t=0} = E(X^k)$$

- Based on the MGF, we can show that linear combinations of independent normal random variables are also normal.

Moment Generating Functions

- MGF of X is $M_X(t) = E\left[e^{tX}\right] = \int e^{tx} f(x)dx$.

Property D Extension: Let $X_1, \ldots, X_n$ be a random sample from a population with moment generating function $M_X(t)$. Then the sample mean has the following mgf:

$$\begin{aligned}
M_{\bar{X}}(t) &= E\left[e^{t\bar{X}}\right] = E\left[e^{t(X_1+\cdots+X_n)/n}\right] \\
&= E\left[e^{tX_1/n} \times \cdots \times e^{tX_n/n}\right] \\
&= E\left[e^{tX_1/n}\right] \times \cdots \times E\left[e^{tX_n/n}\right] \\
&= M_{X_1}(t/n) \times \cdots \times M_{X_n}(t/n) \\
&= [M_X(t/n)]^n
\end{aligned}$$

Distribution of the Mean of Independent Normal Random Variables

- If $X \sim n(\mu, \sigma^2)$ then the mgf of X is:

$$M_X(t) = E\left[e^{tX}\right] = e^{\mu t + \sigma^2 t^2/2}$$

So For $\bar{X}$ we have:

$$\begin{aligned} M_{\bar{X}}(t) &= \left[exp\left(\mu t/n + \sigma^2 (t/n)^2/2\right)\right]^n \\ &= exp\left(n\left(\mu t/n + \sigma^2 (t/n)^2/2\right)\right) \\ &= exp\left(\mu t + (\sigma^2/n) t^2/2\right) \end{aligned}$$

$$\bar{X} \sim n(\mu, \sigma^2/n)$$

Convergence Concepts - Rice Chapter 5

- An important part of probability theory concerns the behavior of sequences of random variables (large sample theory, limit theory, asymptotic theory).
- What can we say about the limiting behavior of sequences of random variables: $X_1, X_2, X_3, \ldots$?

Definition: A sequence of random variables $X_1, X_2, \ldots$ converges in probability to a random variable X if for every $\epsilon > 0$

$$\lim_{n \to \infty} P(|X_n - X| \geq \epsilon) = 0$$

Definition: A sequence of random variables $X_1, X_2, \ldots$ converges in distribution to a random variable X if

$$\lim_{n \to \infty} F_{X_n}(x) = F_X(x)$$

at all points x where $F_X(x)$ is continuous.

Example: Let $X_n \sim \mathrm{normal}(0, 1/n)$

- We expect X_n to concentrate around 0 as $n \to \infty$.
- Note:

$$\sqrt{n}X_n = Z \quad \sim \quad \mathrm{normal}(0,1)$$

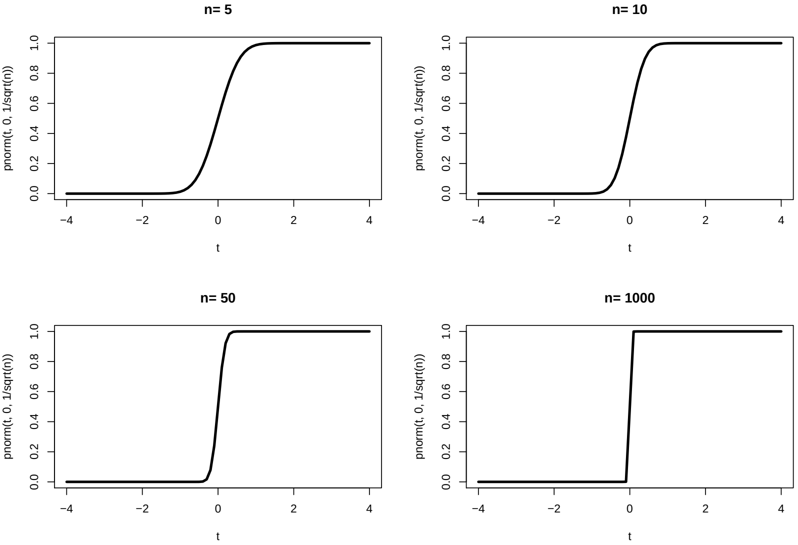
- Let's first consider convergence in distribution:

$$\begin{aligned} F_{X_n}(t) = P(X_n < t) &= P(\sqrt{n}X_n < \sqrt{n}t) \\ &= P(Z < \sqrt{n}t) \end{aligned}$$

  - If $t < 0$ then $P(Z < \sqrt{n}t) \to 0$ as $n \to \infty$.
  - If $t > 0$ then $P(Z < \sqrt{n}t) \to 1$ as $n \to \infty$.

```r
t <- seq(-4, 4, by=0.1)

n.pick <- c(5, 10, 50, 1000)
par(mfrow=c(2,2))
for(n in n.pick){
plot(t, pnorm(t, 0, 1/sqrt(n)), type="l", lwd=3,
     main=paste("n=", n))}
```


n= 5
n= 10
n= 50
n= 1000
pnorm(t, 0, 1/sqrt(n))
t
0.0
0.2
0.4
0.6
0.8
1.0
−4
−2
0
2
4

- We see that we end up with a point mass at 0.

$$X_n \stackrel{D}{\rightarrow} 0$$

- Note: We had the consideration "for all t which F is continuous".

$$F_{x_n}(0) = 1/2 \neq F(0) = 1$$

But we meet the condition, so we are set!

- Now lets consider convergence in probability. For any $\epsilon > 0$ we have:

$$P(|X_n - 0| > \epsilon) \;=\; P(|X_n - 0|^2 > \epsilon^2)$$

- Now using Markov's inequality:

$$\begin{aligned} P(|X_n| > \epsilon) &= P(|X_n - 0|^2 > \epsilon^2) \\ &= P(X_n^2 > \epsilon^2) \\ &\leq \frac{E(X_n^2)}{\epsilon^2} = \frac{1}{n\epsilon^2} \to 0. \end{aligned}$$

As $n \to \infty$, so

$$X_n \xrightarrow{P} 0.$$

Theorem: The following relationships hold:

1. $X_n \stackrel{P}{\to} X$ implies $X_n \stackrel{D}{\to} X$.
2. If $X_n \stackrel{D}{\to} X$ and if $P(X = c) = 1$ (i.e. a point mass) for some real number c, then $X_n \stackrel{P}{\to} X$.

In general the reverses do not hold.

Theorem: Let X_n, X, Y_n, Y be random variables. Let g be a continuous function:

1. If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $X_n + Y_n \xrightarrow{P} X + Y$.
2. If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} c$, then $X_n + Y_n \xrightarrow{D} X + c$.
3. If $X_n \xrightarrow{P} X$ and $Y_n \xrightarrow{P} Y$, then $X_n Y_n \xrightarrow{P} XY$.
4. If $X_n \xrightarrow{D} X$ and $Y_n \xrightarrow{D} c$, then $X_n Y_n \xrightarrow{D} Xc$.
5. If $X_n \xrightarrow{P} X$, then $g(X_n) \xrightarrow{P} g(X)$.
6. If $X_n \xrightarrow{D} X$, then $g(X_n) \xrightarrow{D} g(X)$.

Theorem A: The Weak Law of Large Numbers

- Let $X_1, X_2, \ldots$ be iid random variables.
- $E(X_i) = \mu$ and $V(X_i) = \sigma^2 < \infty$.
- Let $\bar{X}_n = \sum_{i=1}^{n} X_i / n$.
- **Then** $X_n \xrightarrow{P} \mu$.

Proof:

$$
\begin{aligned}
P(|\bar{X}_n - \mu| > \epsilon) &= P((\bar{X}_n - \mu)^2 > \epsilon^2) \\
&\leq \frac{E((\bar{X}_n - \mu)^2)}{\epsilon^2} = \frac{V(\bar{X}_n)}{\epsilon^2} = \frac{\sigma^2}{n\epsilon^2} \to 0
\end{aligned}
$$

as $n \to \infty$.

Central Limit Theorem

Rice Chap 5 Theorem B (extended a bit):

- Let $X_1, X_2, \ldots,$ be a sequence of i.i.d. random variables whose mgfs exist in a neighborhood around 0 (that is, $M_{X_i}(t)$ exists for $|t| < h$ for some positive h).
- Let $E[X_i] = \mu$ and $V[X_i] = \sigma^2 > 0$. Both μ and σ^2 exist because the mgf exists.
- Define $\bar{X} = \frac{1}{n}\sum_{i=1}^{n} X_i$.
- Let $G_n(x)$ denote the cdf of $\frac{(\bar{X}-\mu)}{\sigma/\sqrt{n}}$
- The for any x, $-\infty < x < \infty$, we have:

$$\lim_{n\to\infty} G_n(x) = \int_{-\infty}^{x} \frac{1}{\sqrt{2\pi}} exp(-y^2/2)dy = \mathrm{normal}(0,1)$$

Proof:

- Application mgf for sums of random variables.
- Taylor series expansion of the mgf around 0.
- Use of the properties of mgfs ($E[X]$, $E[X^2]$).
- Remainder of the Taylor series expansion goes to zero.
- $\lim_{n \to \infty} \left(1 + \frac{x}{n}\right)^n = e^x$
- What does the theorem tell us?
    - From very little assumptions we end up with normality.
    - However, for a given sample size we don't know how good the approximation is. $n = 30$?
    - For each situation, if you can computationally check, you may wish to do so.

Proof:

- Define $Y_i = \frac{X_i - \mu}{\sigma}$. Thus $E(Y) = 0$ and $V(Y) = 1$.
- Consider:

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} = \frac{\sqrt{n}(\bar{X}_n - \mu)}{\sigma} = \frac{1}{\sqrt{n}} \sum Y_i$$

- We have a summation, so let's use the MGF:

$$M_{\frac{1}{\sqrt{n}}\sum Y_i}(t) = M_{\sum Y_i}\left(t/\sqrt{n}\right) = \left(M_Y\left(t/\sqrt{n}\right)\right)^n$$

- Now do a Taylor series expansion of the MGF around 0.

$$M_Y\left(t/\sqrt{n}\right) = \sum_{k=0}^{\infty} M_Y^{(k)}(0) \frac{(t/\sqrt{n})^k}{k!}$$

Note: $M_Y^{(k)}(0) = \frac{d^k}{dt^k} M_Y(t)\Big|_{t=0}$

- Recall:

$$e^{tX} = 1 + \frac{tX}{1!} + \frac{(tX)^2}{2!} + \frac{(tX)^3}{3!} + \cdots$$

$$E\left(e^{tX}\right) = 1 + \frac{tE[X]}{1!} + \frac{t^2E[X^2]}{2!} + \frac{t^3E[X^3]}{3!} + \cdots$$

- So by construction of Y we have:
- $M_Y^{(0)}(0) = 1$
- $M_Y^{(1)}(0) = 0$
- $M_Y^{(2)}(0) = 1$

$$
\begin{aligned}
\sum_{k=0}^{\infty} M_Y^{(k)}(0)\frac{(t/\sqrt{n})^k}{k!} &= 1 + 0 + \frac{(t/\sqrt{n})^2}{2!} + M_Y^{(3)}(0)\frac{(t/\sqrt{n})^3}{3!}\cdots \\
&= 1 + \frac{(t/\sqrt{n})^2}{2!} + M_Y^{(3)}(0)\frac{(t/\sqrt{n})^3}{3!}\cdots \\
&= 1 + \frac{(t/\sqrt{n})^2}{2!} + R_Y\left(\frac{t}{\sqrt{n}}\right)
\end{aligned}
$$

- It can be shown for a fixed $t \neq 0$:

$$\lim_{n \to \infty} \frac{R_Y\left(t/\sqrt{n}\right)}{\left(1/\sqrt{n}\right)^2} = \lim_{n \to \infty} nR_Y\left(t/\sqrt{n}\right) = 0$$

- We also note that at $t = 0$ we have:

$$R_Y\left(0/\sqrt{n}\right) = 0$$

$$
\begin{aligned}
\sum_{k=0}^{\infty} M_Y^{(k)}(0)\frac{(t/\sqrt{n})^k}{k!} &= 1 + 0 + \frac{(t/\sqrt{n})^2}{2!} + M_Y^{(3)}(0)\frac{(t/\sqrt{n})^3}{3!}\cdots \\
&= 1 + \frac{(t/\sqrt{n})^2}{2!} + R_Y\left(t/\sqrt{n}\right) \\
&= 1 + \frac{1}{n}\left(\frac{t^2}{2} + nR_Y\left(t/\sqrt{n}\right)\right) \\
&= 1 + \frac{a_n}{n}
\end{aligned}
$$

Math Fact: Let $a_1, a_2, \ldots$ be a sequence of number converging to a, that is $\lim_{n\to\infty} a_n = a$ then

$$\lim_{n\to\infty}\left(1+\frac{a_n}{n}\right)^n = e^a$$

- Now consider:

$$M_{\frac{1}{\sqrt{n}}\sum Y_i}(t) = \left(M_Y\left(t/\sqrt{n}\right)\right)^n = \left(1+\frac{a_n}{n}\right)^n \to exp\left(t^2/2\right)$$

as $n \to \infty$. Since $a_n \to a = t^2/2$. So we have the moment generating function for a standard normal distribution.

CLT Example

1. For i$= 1, \ldots S = 10,000$:
    1.1 Draw a sample of n=10 from an exponential distribution with mean $\beta = 5$.

$$f(x) = \frac{1}{\beta} exp(-x/\beta)$$

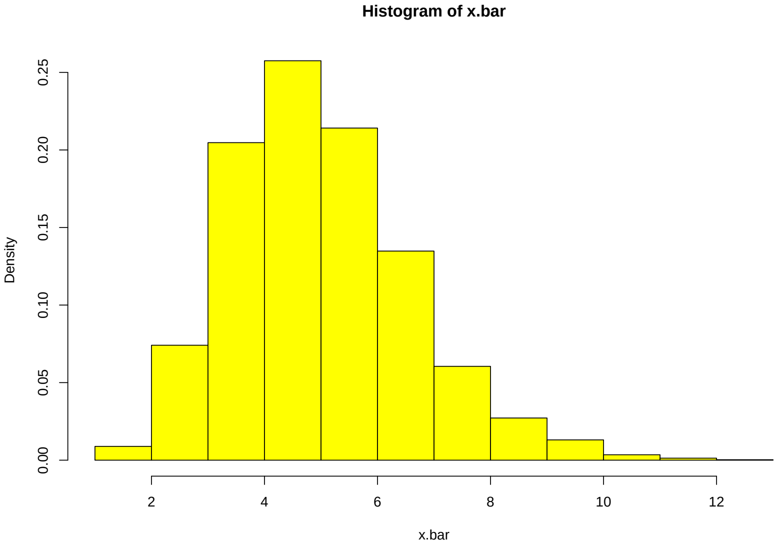
    1.2 Take the mean of $X_1, \ldots, X_n$.
2. Make a histogram of the $\bar{X}_1, \ldots, \bar{X}_S$.

```r
set.seed(1001)

##
S <- 10000
n <- 10
x.bar <- rep(0, S)
for(s in 1:S){
  x.bar[s] <- mean(rexp(n, rate=1/5))
  }
```

- Sample size of $n = 10$ is pretty right skewed!

```
hist(x.bar, prob=TRUE, col="yellow")
```



- Let's try $n = 30$. Still a bit right skewed, but much better.

```r
hist(x.bar, prob=TRUE, col="yellow")
```