moment generating functions

powerful math tricks for dealing with distributions
we won’t do much with it, but mentioned/used in book, so
a very brief introduction:

the \( k^{th} \) moment of r.v. \( X \) is \( \text{E}[X^k] \); M.G.F. is \( M(t) = \text{E}[e^{tX}] \)

\[
e^{tX} = \frac{X^0}{0!} + \frac{X^1}{1!} t^1 + \frac{X^2}{2!} t^2 + \frac{X^3}{3!} t^3 + \ldots
\]

\[
M(t) = \text{E}[e^{tX}] = \text{E}[X^0] \frac{t^0}{0!} + \text{E}[X^1] \frac{t^1}{1!} + \text{E}[X^2] \frac{t^2}{2!} + \text{E}[X^3] \frac{t^3}{3!} + \ldots
\]

\[
\frac{d^2}{dt^2} M(t) = 0 + 0 + \text{E}[X^2] + \text{E}[X^3] \frac{t^1}{1!} + \ldots
\]

\[
\frac{d^2}{dt^2} M(t) \bigg|_0 = \text{E}[X^2]
\]

\[
\frac{d^k}{dt^k} M(t) \bigg|_0 = \text{E}[X^k]
\]
the law of large numbers & the CLT

$$\Pr \left( \lim_{n \to \infty} \left( \frac{X_1 + \cdots + X_n}{n} \right) = \mu \right) = 1$$
weak law of large numbers

i.i.d. (independent, identically distributed) random vars

\( X_1, X_2, X_3, \ldots \)

\( X_i \) has \( \mu = \text{E}[X_i] < \infty \) and \( \sigma^2 = \text{Var}[X_i] \)

Consider the empirical mean:

\[
\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i
\]

The Weak Law of Large Numbers:
For any \( \varepsilon > 0 \), as \( n \to \infty \)

\[
\Pr(|\overline{X} - \mu| > \varepsilon) \longrightarrow 0.
\]
For any $\varepsilon > 0$, as $n \to \infty$

$$\Pr(|\overline{X} - \mu| > \varepsilon) \to 0.$$ 

**Proof:** (assume $\sigma^2 < \infty$)

$$E[\overline{X}] = E\left[\frac{X_1 + \cdots + X_n}{n}\right] = \mu$$

$$\text{Var}[\overline{X}] = \text{Var}\left[\frac{X_1 + \cdots + X_n}{n}\right] = \frac{\sigma^2}{n}$$

By Chebyshev inequality,

$$\Pr(|\overline{X} - \mu| \geq \varepsilon) \leq \frac{\sigma^2}{n\varepsilon^2} \to 0$$
strong law of large numbers

i.i.d. (independent, identically distributed) random vars

\[ X_1, X_2, X_3, \ldots \]

\[ \overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \]

\( X_i \) has \( \mu = \mathbb{E}[X_i] < \infty \)

\[
\Pr \left( \lim_{n \to \infty} \left( \frac{X_1 + \cdots + X_n}{n} \right) = \mu \right) = 1
\]

Strong Law \( \Rightarrow \) Weak Law (but not vice versa)

Strong law implies that for any \( \varepsilon > 0 \), there are only finite number of \( n \) such that the weak law condition \( |\overline{X} - \mu| \geq \varepsilon \) is violated.
diffusion

the law of large numbers

Note: $D_n = E[ | \sum_{1 \leq i \leq n} (X_i - \mu) | ]$ grows with $n$, but $D_n/n \to 0$

Justifies the “frequency” interpretation of probability

“Regression toward the mean”

Gambler’s fallacy: “I’m due for a win!”

“Result will usually be close to the mean”

Many web demos, e.g.
http://stat-www.berkeley.edu/~stark/Java/Html/lln.htm
• X is a normal random variable  \( X \sim N(\mu, \sigma^2) \)

\[
f(x) = \frac{1}{\sigma \sqrt{2\pi}} e^{-\frac{(x-\mu)^2}{2\sigma^2}}
\]

\[
E[X] = \mu \quad \text{Var}[X] = \sigma^2
\]
normal random variable

$X$ is a normal random variable \( X \sim N(\mu, \sigma^2) \)

\( Z \sim N(0,1) \) \( \text{“standard (or unit) normal”} \)

Use \( \Phi(z) \) to denote CDF, i.e.

\[
\Phi(z) = \Pr(Z \leq z) = \int_{-\infty}^{z} \frac{1}{\sqrt{2\pi}} e^{-x^2/2} \, dx
\]

no closed form 😞
the central limit theorem (CLT)

i.i.d. (independent, identically distributed) random vars
\[ X_1, X_2, X_3, \ldots \]

\( X_i \) has \( \mu = E[X_i] \) and \( \sigma^2 = \text{Var}[X_i] \)

As \( n \to 1 \),
\[
\frac{X_1 + X_2 + \cdots + X_n - n\mu}{\sigma \sqrt{n}} \quad \xrightarrow{} \quad N(0, 1)
\]

Restated: As \( n \to \infty \),
\[
\overline{X} = \frac{1}{n} \sum_{i=1}^{n} X_i \sim N \left( \mu, \frac{\sigma^2}{n} \right)
\]
CLT in the real world

CLT is the reason many things appear normally distributed
Many quantities = sums of (roughly) independent random vars

Exam scores: sums of individual problems
People’s heights: sum of many genetic & environmental factors
Measurements: sums of various small instrument errors

...
Human height is approximately normal.

Why might that be true?

R.A. Fisher (1918) noted it would follow from CLT if height were the sum of many independent random effects, e.g. many genetic factors (plus some environmental ones like diet). I.e., suggested part of mechanism by looking at shape of the curve.
CLT convergence
Chernoff bound was CLT (for binomial) in disguise

• Suppose $X \sim Bin(n, p)$
• $\mu = E[X] = pn$

• Chernoff bound:

For any $\delta$ with $0 < \delta < 1$,

$$P(X > (1 + \delta)\mu) \leq e^{-\frac{\delta^2 \mu}{2}}$$

$$P(X < (1 - \delta)\mu) \leq e^{-\frac{\delta^2 \mu}{3}}$$
• Roll 10 6-sided dice
• \( X = \text{total value of all 10 dice} \)
• Win if: \( X \leq 25 \) or \( X \geq 45 \)
• Roll…

\[
\begin{align*}
E[X] &= 10E[X_i] = 10(3.5) = 35 \\
\text{Var}(X) &= 10 \text{Var}(X_i) = 10 \frac{35}{12} = \frac{350}{12} \\
1 - P(25.5 \leq X \leq 44.5) &= 1 - P\left(\frac{25.5 - 35}{\sqrt{350}/12} \leq \frac{X - 35}{\sqrt{350}/12} \leq \frac{44.5 - 35}{\sqrt{350}/12}\right) \\
&\approx 1 - (2\Phi(1.76) - 1) \approx 2(1 - 0.9608) = 0.0784
\end{align*}
\]