

Can show that $B(\alpha, \beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$. (282)

(α, β) jointly control

the shape of the Beta(α, β) dist.

$$Z \sim \text{Beta}(\alpha, \beta) \quad F_Z(t) = 1 + \sum_{k=1}^{\infty} \left(\prod_{r=0}^{k-1} \frac{\alpha+r}{\alpha+\beta+r} \right) \frac{t^k}{k!}$$

$$E(Z) = \frac{\alpha}{\alpha+\beta}$$

$$V(Z) = \left(\frac{\alpha}{\alpha+\beta} \right) \left(\frac{\beta}{\alpha+\beta} \right) \left(\frac{1}{\alpha+\beta+1} \right)$$

Example $n=220$ grand jurors chosen from (19 Aug 16)

(Castaneda v. Partida continued) eligible population of Hidalgo County, Texas, which was 79.1% Mexican-American, but only $s=100$

selected grand jurors were Mexican-American; summarize the information in a Bayesian fashion about evidence of discrimination.

Data) $S = \# \text{ Mexican-American } \xrightarrow{\text{chosen}} \text{ in jury}$ selection of $n = 220$ people 283

Unknown) $\theta = \text{actual probability of an eligible Mexican-American person being chosen}$
 $(0 < \theta < 1)$

Model) $(S' | \theta) \sim \text{Binomial}(n, \theta),$

$$\text{i.e., } f_{S|\theta}(s|\theta) = P(S=s|\theta) = \binom{n}{s} \theta^s (1-\theta)^{n-s}.$$

Bayesian approach) ① Information internal to data set about θ summarized by the likelihood (un-normalized) density,

defined to be
$$l(\theta | s) = c P(S=s | \theta),$$

can arbitrary positive constant — think of $P(S=s | \theta)$ as a function of θ for fixed s .

$$\text{Here } l(\theta | s) = c \binom{n}{s} \theta^s (1-\theta)^{n-s} \quad \begin{array}{l} \text{can be } 284 \\ \text{absorbed} \end{array}$$

$$= c \theta^s (1-\theta)^{n-s} \quad \begin{array}{l} \text{into } c \text{ since} \\ \text{does not depend} \\ \text{on } \theta \end{array}$$

(2) Information external

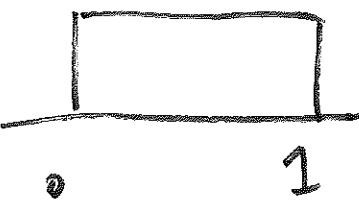
To dataset about θ summarized

by the prior density $f_{\Theta}(s)$. Here are some

possibilities for the prior, depending on information

your knowledge base:

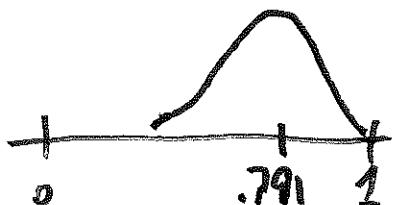
(a) neutral $\Theta \sim \text{Uniform}(0,1)$



this dist. embodies the

information { θ could be anywhere
between 0 and 1, with no value favored}

(b) cut the district attorney some slack prior



This prior gives the DA the benefit
of the doubt

When you're uncertain about what prior 285 to use, write down all the reasonable priors & do a sensitivity analysis (use each prior one by one & see if answer is the same)

③ Combine internal & external information

with

Bayes' theorem

Theorem

Here

$$f_{\text{IS}}(\theta | s) = c \cdot f_{\text{prior}}(\theta) \cdot f_{\text{likelihood}}(s | \theta)$$

$$\text{posterior information} = (\text{normalizing constant}) \cdot (\text{prior information}) \cdot (\text{likelihood information})$$

$$f_{\text{IS}}(\theta | s) = c f_{\theta}(s) \theta^s (1-\theta)^{n-s}$$

Rev. Bayes himself waited back in 1760

that if you take $f_{\theta}(x) = c \theta^{\text{prior power}} (1-\theta)^{\text{prior power}}$
 then the product of 2 such densities is
 another such density, meaning that the
 posterior would have the same form as
 the prior & likelihood, making calculating

easier

Moreover, we already know the

name of densities that look like $\theta^{\text{prior power}} (1-\theta)^{\text{prior power}}$:

the $(X \sim \text{Beta}(\alpha, \beta) \quad (\alpha > 0, \beta > 0)) \rightarrow$

Beta

distributions

$$f_X(x) = c \theta^{\alpha-1} (1-\theta)^{\beta-1}$$

so let's take $f_{\theta}(x) = c \theta^{\alpha-1} (1-\theta)^{\beta-1}$

in the law suit case study; then

$$f_{\theta|S}(\theta|s) = c [\theta^{\alpha-1} (1-\theta)^{\beta-1}] [\theta^s (1-\theta)^{n-s}]$$

$$= c \theta^{(d+s)-1} (1-\theta)^{(\beta+n-s)-1} = \text{Beta}(\alpha+s, \beta+n-s)$$

(287)

So the prior-to-posterior

updating looks like this:

"Beta dist. is
conjugate to the
Binomial likelihood"

$$\begin{aligned} \theta &\sim \text{Beta}(\alpha, \beta) \\ (s'|\theta) &\sim \text{Binomial}(n, \theta) \end{aligned} \quad \left. \begin{array}{l} \theta|s \sim \text{Beta}(\alpha+s, \\ \beta+n-s) \end{array} \right\}$$

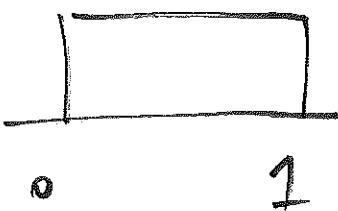
$$s = 100$$

$$n = 220$$

How choose (α, β) ?

(a) Neutral prior

$$\text{But } \text{Uniform}(0,1) = \theta^{1-1} (1-\theta)^{1-1}$$



$$\text{So } \theta \sim \text{Uniform}(0,1) \leftrightarrow \theta \sim \text{Beta}(1,1)$$

(b) cut DA slack prior

There's an extremely useful thing that happens with conjugate priors:

$\begin{bmatrix} 0 \\ \vdots \\ 1 \\ \vdots \\ 1 \end{bmatrix} \cdot \begin{bmatrix} \alpha \\ \beta \end{bmatrix}$

mean $\frac{\alpha}{\alpha+\beta}$

$\left. \begin{array}{l} \text{prior} \\ \text{effective} \\ \text{sample} \\ \text{size } (\alpha+\beta) \end{array} \right\}$

Beta prior distribution acts like a dataset with $\alpha=15$ & $\beta=5$

with the property that

$\begin{bmatrix} 0 \\ \vdots \\ 0 \\ \vdots \\ 1 \end{bmatrix} \cdot \begin{bmatrix} n-s \\ s \end{bmatrix}$

data set
sample size n

$$\text{mean } \bar{y} = \frac{s}{n}$$

if you do a Bayesian analysis with the Beta(α, β) prior and I do a frequentist

analysis on the dataset with $(\alpha+s)$ 1s and $(\beta+n-s)$ 0s formed by merging the prior & sample datasets, we'll get the same results.

(b) cut
the DA
slack
prior

mean of $\text{Beta}(\alpha, \beta)$ dist. is $\frac{\alpha}{\alpha + \beta}$
 $\frac{\alpha}{\alpha + \beta}$; set this equal to 0.791

Suppose I want to put in information equivalent to a prior sample size $\frac{1}{10}$ or λ_{ij} as the data sample size (say); set

$$(\alpha + \beta) = \frac{1}{10} n = 22$$

$$n = 22$$

$$s = 10$$

Solve: $\begin{cases} \alpha = 17.4 \\ \beta = 4.6 \end{cases}$

likelihood is

$$C \theta^s (1-\theta)^{n-s} = C \theta^{(s+1)-1} (1-\theta)^{(n-s+1)-1}$$

(a) Neutral prior:

$$\text{Beta}(1, 1)$$

prior sample size 2

= Beta($s+1$, $n-s+1$) dist

(10)

(21)

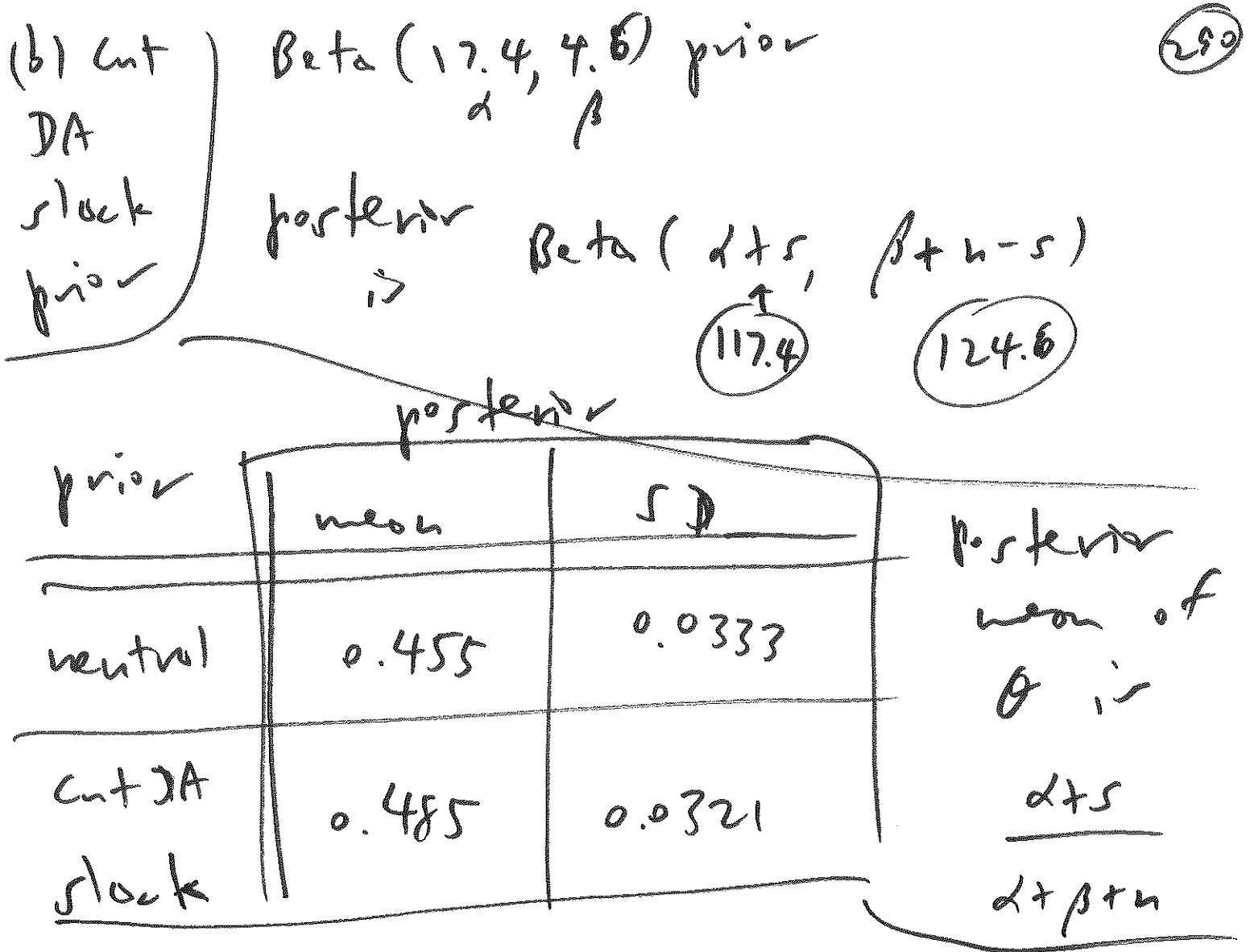
posterior

is $\text{Beta}(\alpha+s, \beta+n-s)$

(10)

(21)

(same as like likelihood)



Posterior SD is

$$\sqrt{\left(\frac{\alpha + s}{\alpha + \beta + n}\right)\left(\frac{\beta + n - s}{\alpha + \beta + n}\right)\left(\frac{1}{\alpha + \beta + n + 1}\right)}$$

The no-discrimination rate of 0.791 is

$$\frac{0.791 - 0.455}{0.0333} = 12.6 \text{ SDs away from posterior expectation}$$

under the neutral prior and

(29)

$$\frac{0.791 - 0.485}{0.321} = 9.5 \text{ posterior SDs}$$

away from posterior expectation under
the uniform flat prior; Q.E.D.

Multinomial

Distributions

You're contemplating a
population that contains
elements of $k \geq 2$ types

(e.g., {Democrat, Republican, Libertarian,
Independent, Green}).

Suppose the proportion
of elements of type i is \hat{p}_i

with $\sum_{i=1}^k p_i = 1$; $\mathbf{p} = (p_1, \dots, p_k)$.

You take an IID sample of size n from this pop.; $\bar{X}_i = \# \text{elements of type } i \text{ in your sample}; \sum_{i=1}^k \bar{X}_i = n$. (292)

Can show that the vector $\bar{\mathbf{X}} = (\bar{X}_1, \dots, \bar{X}_k)$

has PF

$$f_{\bar{\mathbf{X}}|n,p}(x|n, p) = \begin{cases} \binom{n}{x_1, \dots, x_k} p_1^{x_1} \cdots p_k^{x_k} & \text{if } \sum_{i=1}^k x_i = n \\ 0 & \text{else} \end{cases}$$

where

$$\left(\binom{n}{x_1, \dots, x_k} \right) \stackrel{\text{def}}{=} \frac{n!}{x_1! x_2! \cdots x_k!}$$

$\left(\sum_{i=1}^k p_i = 1 \right)$

$\left(\binom{n}{x_1, \dots, x_k} \right) \stackrel{\text{def}}{=} \frac{n!}{x_1! x_2! \cdots x_k!}$ is the multinomial coefficient

This is called the multinomial (n, p) distribution.

$$E(\Xi_i) = np_i \quad V(\Xi_i) = np_i(1-p_i)$$

(just like binomial)) But now something new:

$$C(\Xi_i, \Xi_j) = -np_i p_j$$

negatively correlated

because $\sum_{i=1}^k \Xi_i = n$

Bivariate normal dist. | Can build a 2-dimensional (bivariate) version of the Dist. bivariate dist. as follows:

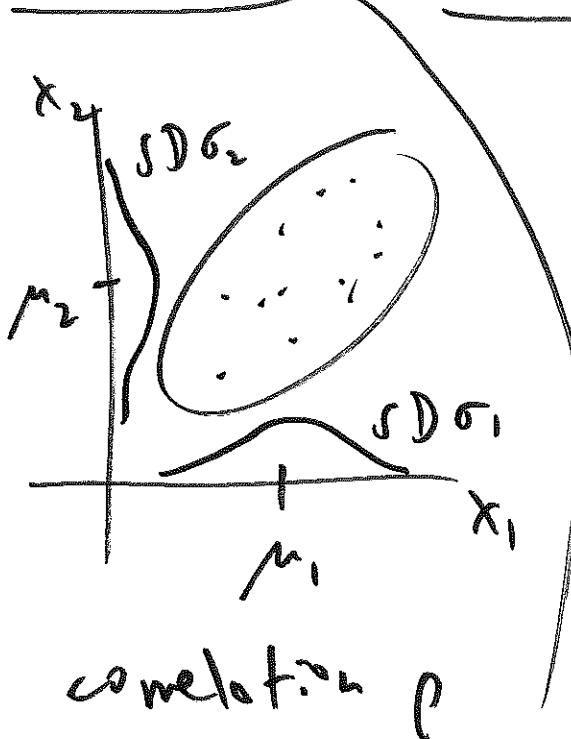
$$\zeta_1, \zeta_2 \stackrel{\text{ID}}{\sim} N(\mu, 1)$$

Specify 5 parameters:

$$-\infty < \mu_1 < \infty \quad 0 < \sigma_1 < \infty$$

$$-\infty < \mu_2 < \infty \quad 0 < \sigma_2 < \infty$$

$$-1 < \rho < 1$$



Now build (\bar{X}_1, \bar{X}_2) with the transformation $\bar{X}_1 = \mu_1 + \sigma_1 \xi_1$

$$\bar{X}_2 = \sigma_2 \left[\rho \xi_1 + \sqrt{1-\rho^2} \xi_2 \right] + \mu_2$$

The joint PDF of $\underline{X} = (\bar{X}_1, \bar{X}_2)$ is

then $f_{\bar{X}_1, \bar{X}_2}(x_1, x_2) = \frac{1}{2\pi\sqrt{1-\rho^2}\sigma_1\sigma_2} \cdot \exp \left\{ -\frac{1}{2(1-\rho^2)} \left[\left(\frac{x_1 - \mu_1}{\sigma_1} \right)^2 - 2\rho \left(\frac{x_1 - \mu_1}{\sigma_1} \right) \left(\frac{x_2 - \mu_2}{\sigma_2} \right) \right. \right.$

$$\left. \left. + \left(\frac{x_2 - \mu_2}{\sigma_2} \right)^2 \right] \right\}$$

~~standard units~~

This is the Bivariate Normal ($\mu_1, \mu_2, \sigma_1, \sigma_2, \rho$) dist.

Easy to show that $E(\bar{X}_1) = \mu_1$, (295)

$E(\bar{X}_2) = \mu_2$, $V(\bar{X}_1) = \sigma_1^2$, $V(\bar{X}_2) = \sigma_2^2$,

$$\rho(\bar{X}_1, \bar{X}_2) = \rho \cdot \boxed{\text{Consequence of } \rho_{ij} \text{ def.}}$$

① $(\bar{X}_1, \bar{X}_2) \sim \text{Bivariate Normal} \rightarrow$

$$\left(\begin{array}{c} \bar{X}_1, \bar{X}_2 \\ \text{: independent} \end{array} \right) \leftrightarrow \left(\begin{array}{c} \bar{X}_1, \bar{X}_2 \\ \text{uncorrelated} \end{array} \right)$$

we already knew the \rightarrow direction is general; what's new here is that correlation 0 implies independence

if $(\bar{X}_1, \bar{X}_2) \sim \text{Bivariate Normal}$.

② $(\bar{X}_1, \bar{X}_2) \sim \text{Bivariate Normal}(\mu_1, \mu_2, \sigma_1^2, \sigma_2^2, \rho)$ (296)

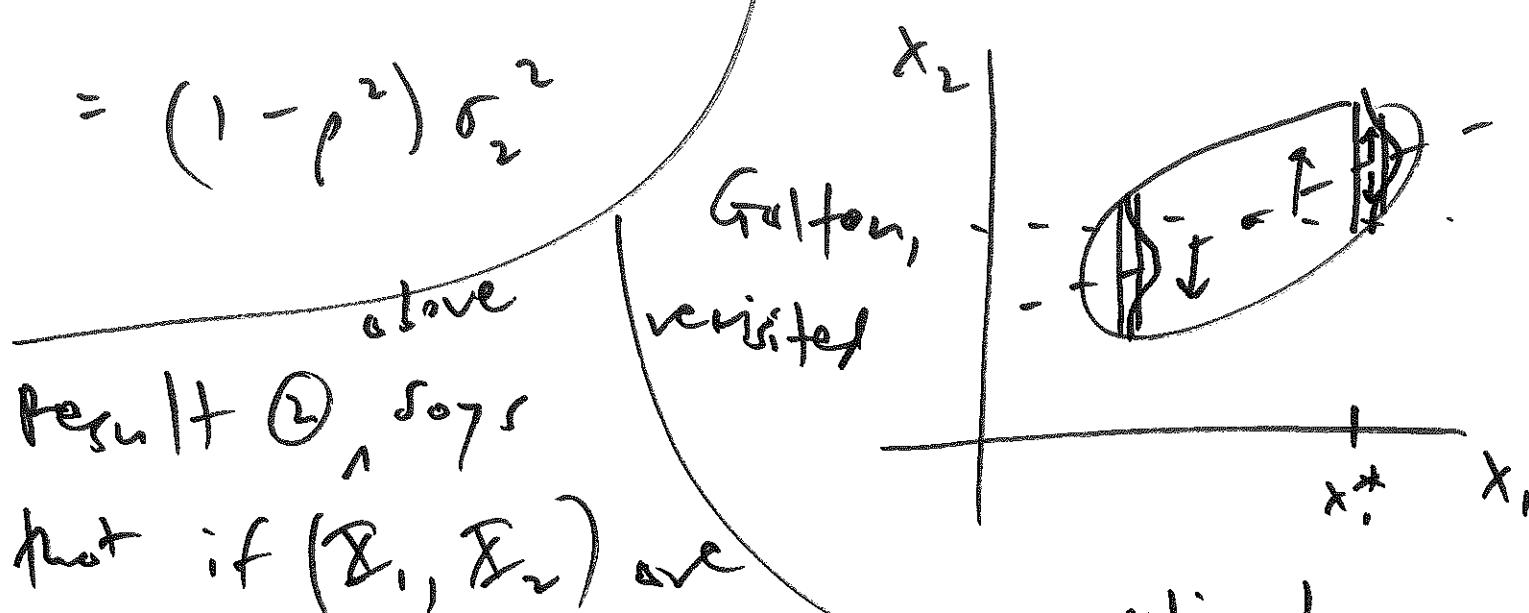
+ conditional distribution of \bar{X}_2

given that $\bar{X}_1 = x_1$ is (univariate) normal with mean $E(\bar{X}_2 | x_1) =$

$$\text{and variance } V(\bar{X}_2 | x_1) =$$

$$= (1 - \rho^2) \sigma_2^2$$

$$= \frac{\sigma_2^2}{\sigma_1^2} (x_1 - \mu_1)^2$$



result ③ says

that if (\bar{X}_1, \bar{X}_2) are

Bivariate Normal then the distributions of \bar{X}_2 given $\bar{X} = x^*$ are all of the vertical strips are also normal

And the means of all these normal distributions in the vertical strips are connected together by Galton's

regression
line

$$\hat{x}_2 = \mu_2 + \left(\frac{\sigma_2}{\sigma_1} (x_1 - \mu_1) \right).$$

This line has slope $\beta_1 = \frac{\sigma_2}{\sigma_1}$ and "y-intercept"

$$\beta_0 = \mu_2 - \beta_1 \mu_1$$

Moreover,

$$\hat{x}_2 = \beta_0 + \beta_1 x_1$$

we can now quantify an earlier insight:

ignore x_1 , predict $(\hat{x}_2)_{x_1=0} = \mu_2 = E(\bar{X}_2)$

(root mean squared error) (RMSE) of this prediction is

$$\sqrt{V(\bar{X}_2)} = \sigma_2$$

use x_1
to predict
 x_2

$$\text{pred.2t } (\hat{x}_2)_{\text{use}} = E(\bar{X}_2 | \bar{X}_1 = x_1)$$

$$= \mu_2 + \frac{\rho \sigma_2}{\sigma_1} (x_1 - \mu_1)$$

part of the

prediction is $\sqrt{V(\bar{X}_2 | x_1)} = \sigma_2 \sqrt{1 - \rho^2}$

Since $-1 < \rho < 1$, $\sigma_2 \sqrt{1 - \rho^2} \leq \sigma_2$

with equality only when $\rho = 0$.

③ $(\bar{X}_1, \bar{X}_2) \sim \text{Bivariate Normal}(\mu_1, \mu_2, \sigma_1, \sigma_2, \rho)$

$$Y = a_1 \bar{X}_1 + a_2 \bar{X}_2 + b, \quad (a_1, a_2, b) \text{ arbitrary constants}$$

$$\rightarrow Y \sim N(a_1 \mu_1 + a_2 \mu_2 + b, a_1^2 \sigma_1^2 + a_2^2 \sigma_2^2 + 2a_1 a_2 \rho \sigma_1 \sigma_2)$$

large
Random
Samples
(DS ch. 6)

You know n IID random
sample X_1, \dots, X_n from a population,
with the goal of estimating the
population mean $\mu = E(X_i)$.

We've already seen that, from a most
mean squared error point of view, the
sample mean $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ is the best
you can do (in the absence of prior
information).

It would be nice if
 \bar{X}_n approached the
right answer μ as n increases; how
to quantify that idea?

Two
inequalities
that
help

300

Markov inequality

Suppose X is a non-negative r.v., i.e. $P(X \geq 0) = 1$

Then for all

$$\text{real } t > 0, P(X \geq t) \leq \frac{E(X)}{t} \quad (*)$$

(Attributed to Andrey Markov (1856-1922),
 a Russian mathematician who did pioneering
 work on stochastic processes)

(*) Says that, if $E(X)$ is fixed,
 you can't move more & more
 probability out into the
 right tail beyond a
 certain point.

to place
↓
Bartels → Lobachevsky → Brashman → Chebyshev

Droper
↑
Lebesgue
↑
Neyman
↑
Sierpiński
↑
Voronoy
↑
Markov

ex. $E(\bar{X}) = 1$, \bar{X} non-negative \rightarrow (30)

$$P(\bar{X} \geq 100) \leq \frac{1}{100}$$

The inequality is

sharp, meaning that the upper bound

$\frac{E(\bar{X})}{t}$ on $P(\bar{X} \geq t)$ is achievable, \oplus

ex. $E(\bar{X}) = 1$, \bar{X} -nonnegative \rightarrow

put probability 0.99 on $\bar{X} = 0$ and
0.01 on $\bar{X} = 100$

\oplus but most of the time (i.e., for most distributions) it's a crude upper bound.

Can apply Markov inequality to the
rv. $\bar{X}^2 = (\bar{X} - E(\bar{X}))^2$ to get

Chebyshev) \mathbb{X} w. with $V(\mathbb{X})$ existing 302
Inequality + for every $t \geq 0$,

$$P(|\mathbb{X} - E(\mathbb{X})| \geq t) \leq \frac{V(\mathbb{X})}{t^2}$$

(attributed
to

Pafnuty Chebyshev (1821 - 1894), also a Russian mathematician, one of whose Ph.D. students was Markov

Ex.

$$\begin{aligned} E(\mathbb{X}) &= \mu \\ V(\mathbb{X}) &= \sigma^2 \end{aligned}$$

Chebyshev says $P\left[\left|\frac{\mathbb{X}-\mu}{\sigma}\right| \geq 3\right] \leq \frac{1}{3^2} = \frac{1}{9}$,

so no more than $\frac{1}{9} = 11\%$ of the probability in any distribution can be more than 3 SDs away from the mean (for Normal dist. this prob. is 0.3%)

This upper bound is also sharp, but for most distributions it's (also) crude (as with the Markov bound). 303

Back to \bar{X}_n

$X_i \stackrel{\text{iid}}{\sim}$ some dist. with mean $E(X_i) = \mu$
 $(i=1, \dots, n)$ and variance $V(X_i) = \sigma^2$

Then we already showed that if $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$

then $E(\bar{X}_n) = \mu$ for all $n = 1, 2, \dots$

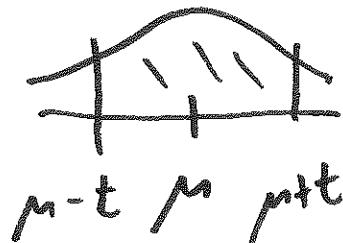
and $V(\bar{X}_n) = \frac{\sigma^2}{n}$. Chebyshev then

gives $P(|\bar{X}_n - \mu| \geq t) \leq \frac{\sigma^2}{nt^2}$ for all $t > 0$

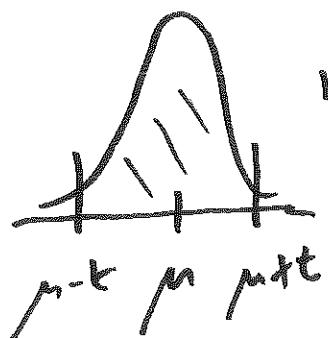
This can be

rewritten $P(|\bar{X}_n - \mu| < t) \geq 1 - \frac{\sigma^2}{nt^2}$

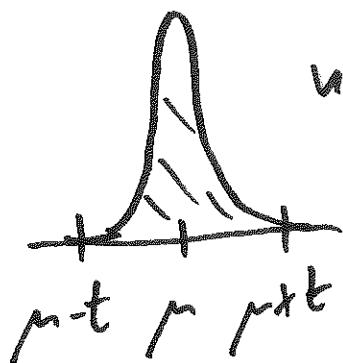
PDF of \bar{X}_n $n=1$



$n=10$



$n=100$



⋮

This suggests a way 304 to quantify how close a r.v. like \bar{X}_n is to a constant like μ :

Def. A sequence $\bar{Z}_1, \bar{Z}_2, \dots$ of r.v. is said to converge in probability to a constant b if

for all $\epsilon > 0$, $\lim_{n \rightarrow \infty} P(|\bar{Z}_n - b| < \epsilon) = 1$;

This is denoted $\bar{Z}_n \xrightarrow{P} b$.

This is immediate consequence of Chebyshev & this definition is

(weak) law of large numbers 305
 $X_i \stackrel{\text{IID}}{\sim} \text{dist. with mean } \mu \text{ and variance } \sigma^2 < \infty$, $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$
 $\bar{X}_n \xrightarrow{P} \mu$ This result has
 the Italian mathematician
 a long history: Gerolamo Cardano (1501-1576)
 asserted it without proof; Jacob Bernoulli (1654-1705)
 proved it for ~~$(X_i | \theta)$~~ $\stackrel{\text{IID}}{\sim}$ Bernoulli(θ)
 (it took him 20 years to find the correct
 proof, published posthumously in 1713;
 Bernoulli thought that this theorem proved
 the existence of God); Simeon Denis Poisson
 named it the law of large numbers in
 1837. Corollary If $Z_n \xrightarrow{P} b$ and $g(z)$
 is continuous at $z=b$ then $g(Z_n) \xrightarrow{P} g(b)$.

Central
Limit
Theorem

Example

$$X_i \stackrel{\text{IID}}{\sim} N(\mu, \sigma^2), \sigma < \infty$$

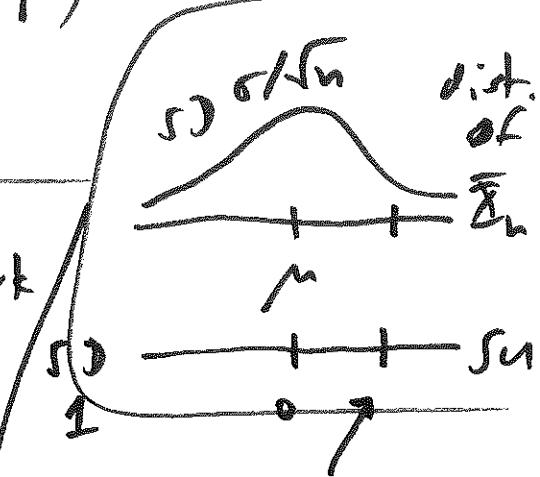
(306)

we know

that $\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i$ has mean μ ,

variance $\frac{\sigma^2}{n}$ and is normally distributed,

so that $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$ for all $n=1, 2, \dots$



A:
Does something like this work
for other choices of f

$$X_i \stackrel{\text{IID}}{\sim} ?$$

? **A:** Yes: most famous result in all of probability.

$$\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}}$$

Central
Limit
Theorem

$X_i \stackrel{\text{IID}}{\sim}$ **any** dist. with mean μ and finite variance $\sigma < \sigma^2 < \infty$,

$\bar{X}_n = \frac{1}{n} \sum_{i=1}^n X_i \rightarrow$ for large n $\frac{\bar{X}_n - \mu}{\sigma/\sqrt{n}} \sim N(0, 1)$

Careful statement Def. $\Sigma_1, \Sigma_2, \dots$ a sequence (3.0)

of rv.; let F_n be the CDF of Σ_n

+ if there exists a CDF F^* such

that $\lim_{n \rightarrow \infty} F_n(x) = F^*(x)$, ^{for} all x at

which $F^*(x)$ is continuous, then

people say that $\Sigma_n \xrightarrow{D} F^*$ (^{"In} convergence
^{in distribution}
to F^* ")

CLT $\Sigma_i \stackrel{\text{IID}}{\sim}$ ^{any} dist. with mean μ
and variance $0 < \sigma^2 < \infty$, $\bar{\Sigma}_n = \frac{1}{n} \sum_{i=1}^n \Sigma_i$

$\sqrt{n} \left(\frac{\bar{\Sigma}_n - \mu}{\sigma} \right) \xrightarrow{D} N(0, 1)$. Re
CLT

also has a long history: it was

first demonstrated for $\mathbb{E}_i \sim$ ^{IID} Bernoulli(θ)⁽³⁾
by the French/British mathematician
Abraham de Moivre (1667 - 1754) in
1733; almost forgotten until revived by
the French mathematician Pierre-Simon de
Laplace (1749 - 1827) in 1812; almost
forgotten again until 1901, when the
Russian mathematician Aleksandr Lyapunov
gave a more general proof; more general
proof provided by J.W. Lindeberg (Finnish
mathematician (1876 - 1932)) and independently
by Paul Lévy (French mathematician (1886 -
1971)) in the early 1920s. CLT name due to
~~Hungarian-American mathematician (1887-1985) George Pólya in 1920~~

Example] Contaminated water supply: (309)

Σ = organic concentration

Σ = lead concentration
(same units) $\left(\frac{\text{b.Pb}}{\text{L.O}}\right)$

Interest focuses

$$\text{as } R = \frac{\Sigma}{\Sigma + \Sigma}$$

(proportion of contamination due to lead)

$E(R) = E\left(\frac{\Sigma}{\Sigma + \Sigma}\right)$ difficult to calculate.

simulation approach] Randomly sample n pairs (Σ_i, Σ_j) from the joint PDF

• $f(\Sigma, \Sigma)$, calculate $R_i = \frac{\Sigma_i}{\Sigma_i + \Sigma_i}$ and

$$\bar{R}_n = \frac{1}{n} \sum_{i=1}^n R_i \leftarrow \text{seed Monte Carlo}$$

(Simulation) estimate of $E(R)$.

(d) How big does \bar{R}_n need to be to achieve $\underline{\text{desired}}$ accuracy target? (310)

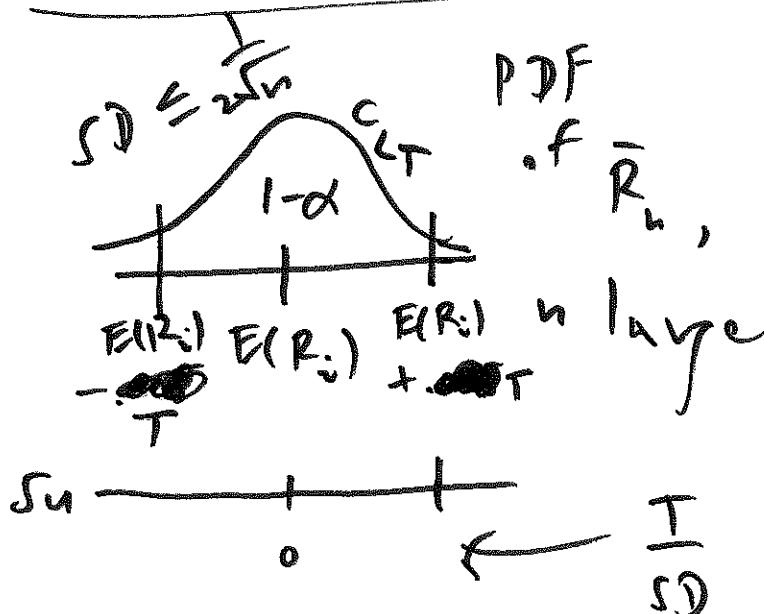
$$|R_i| = \left| \frac{I_i}{\sum I_i} \right| \leq 1; \text{ we show that}$$

$$\text{as a result } V(R_i) \leq \frac{1}{4}. \quad \text{CLT}$$

Says that dist. of \bar{R}_n will be close

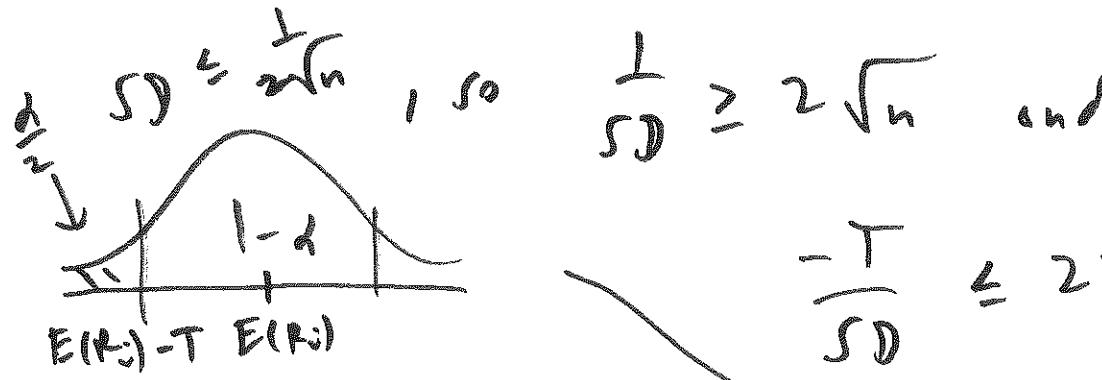
to Normal for large n , with mean $E(\bar{R}_n)$

and Variance $\frac{V(\bar{R}_n)}{n} \leq \frac{1}{4n}$ Suppose we want \bar{R}_n to



differ from $E(\bar{R}_n)$ by no more than some tolerance T with probability at least $(1-\alpha)$...

(311)



$$\frac{-T}{S_D} \leq 2T\sqrt{n}$$

$$I^{-1}\left(\frac{\alpha}{2}\right) = \frac{[E(R_i) - T] - E(R_i)}{S_D} = \frac{-T}{S_D} \leq 2T\sqrt{n}$$

from which $n \geq \left[\frac{I^{-1}\left(\frac{\alpha}{2}\right)}{2T} \right]^2$

For instance

set

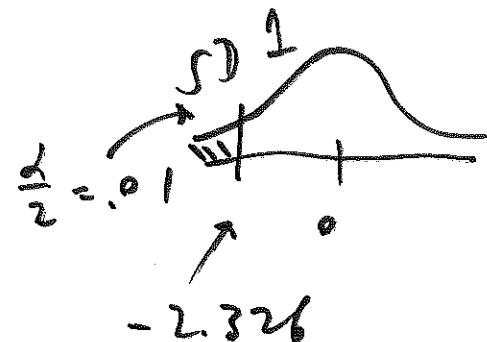
$$T = 0.005$$

($\frac{1}{2}$ of 12)

and $\alpha = .02$ to get

$$n \geq \left[\frac{-2.326}{2(0.005)} \right]^2 \approx 54,119$$

simulation replications



needed Case study: Escalators
in the London Underground ($\frac{22}{18}$)