Last time: $X_1, \dots, X_n$ i.i.d. coin toss with θ chance of Heads

$P(X_1, X_2, \dots, X_n \quad s \text{ Heads}, \ n-s \text{ Tails}) = \underbrace{\int \binom{n}{s} \theta^s (1-\theta)^{n-s} d\theta}_{\text{law of total probability applied to value of } \theta} = \frac{1}{n+1}$

$\parallel$

Z

$\pi_1(\theta) = \frac{1}{Z} \underbrace{P(s \text{ Heads}, \ n-s \text{ Tails} \mid \theta)}_{\binom{n}{s}\theta^s(1-\theta)^{n-s}} \underbrace{\pi_0(\theta)}_{\text{unif}(0,1)} \quad = 1_{(0,1)}(\theta)$

Goal now: compare this integral and get $\frac{1}{n+1}$

$Y = \#$ Heads among $X_1, X_2, \dots, X_n$

This says we should get $P(Y=s) = \frac{1}{n+1}$ for $s = 0, 1, \dots, n$.

i.e. want to show Y is a uniform RV on $\underbrace{\{0, 1, \dots, n\}}_{n+1 \text{ possible values.}}$

$U_1, U_2, \dots, U_n$ be i.i.d. unif$(0,1)$ RV's and jointly independent from θ (remarks: $\theta \sim$ unif$(0,1)$ RV)

$X_i = 1_{U_i \le \theta} = \begin{cases} 1 & U_i \le \theta \\ 0 & U_i > \theta \end{cases}$

$Y = X_1 + X_2 + \dots + X_n$ $\qquad P(1_{U_i \le \theta} = 1) = P(U_i \le \theta) = \theta$ (conditional on θ)

$= 1_{U_1 \le \theta} + \dots + 1_{U_n \le \theta}$

$= \#\{U_i : U_i \le \theta\}$

$Y = \#\{U_i \le \theta\}$

$Y = s$

$\overbrace{U_1, U_2, \dots, U_n, \ \theta = \text{"}U_0\text{"}}^{n+1}$ i.i.d. unif$(0,1)$ RV

$\{Y = s\} = \{\theta > s \text{ of } U\text{'s and } \theta < n-s \text{ of } U\text{'s}\}$

$= \{\theta \text{ is rank } s+1$

$U_1, U_2, \dots, U_n, \theta$

Rank, order from smallest to biggest

$U_0, U_1, \dots, U_n$

$U_{(0)} < U_{(1)} < \dots < U_{(n)}$.

$\uparrow$

$\theta = U_0$

- By symmetry of the U's ($U_0 = \theta$) all ranks are equivalently likely.

$$\left\{\begin{array}{l} = P(U_0 = \theta \text{ has rank } s+1) = \frac{1}{n+1} \\ = P(U_1 \text{ has rank } s+1) = \frac{1}{n+1} \\ = P(U_2 \text{ has rank } s+1) = \frac{1}{n+1} \\ \vdots \\ = P(U_n \text{ has rank } s+1) = \frac{1}{n+1} \end{array}\right.$$

symmetry among $U_1, U_2, \dots, U_n$

add up to 1

Bayesian Estimators

$\pi_0 = \xrightarrow[\substack{s \text{ successes} \\ n-s \text{ failures}}]{\text{observe data}} \pi_1$

$\theta \sim Beta(\alpha, \beta)$ $\quad\quad$ $\theta \sim Beta(\alpha + s, \beta + n - s)$

Get a point estimator $\hat{\theta}$

- choose $\hat{\theta}_B = \mathbb{E}_{\pi_1}(\theta)$ posterior mean of θ
- $\hat{\theta}_B$ minimizes the mean-squared error (MSE) under posterior π_1

$MSE = \mathbb{E}_{\pi_1}(\hat{\theta} - \theta)^2 = Var_{\pi_1}(\theta)$ when $\hat{\theta} = \mathbb{E}_{\pi_1}\theta$

(In general, the mean of π meaning $\mathbb{E}_\pi(\hat{\theta} - \theta)^2$) HW!

$X = X_{n+1}$ independent, new observation Bernoulli (θ)

Lemma $P_{\pi_1}(X = \text{success}) = \hat{\theta}_B := \mathbb{E}_{\pi_1}\theta$

Proof: $P_{\pi_1}(X = \text{success}) = \int P_{\pi_1}(X = \text{success} \mid \theta)\, \pi_1(\theta)\, d\theta$

$= \int \theta\, \pi_1(\theta)\, d\theta$

$= \mathbb{E}_{\pi_1}\theta$

Q: what's $\mathbb{E}X$ for $X \sim Beta(\alpha, \beta)$?

$\mathbb{E}X = \int \theta\, f_X(\theta)\, d\theta$

$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \theta \cdot \theta^{\alpha-1}(1-\theta)^{\beta-1}\, d\theta$

$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 \theta^{\alpha+1-1}(1-\theta)^{\beta-1}\, d\theta$

Beta $(\alpha+1, \beta)$

$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}$ $\quad\quad$ Z for $Beta(\alpha, \beta) = \frac{1}{\frac{\Gamma(\alpha, \beta+1)}{\Gamma(\alpha+1)\Gamma(\beta)}}$

// Z for $Beta(\alpha, \beta)$

$= \frac{\cancel{\Gamma(\alpha+\beta)}}{\cancel{\Gamma(\alpha)}} \cdot \frac{\Gamma(\alpha+1)}{\Gamma(\alpha+\beta+1)}$ $\quad$ $\alpha\cancel{\Gamma(\alpha)}$ $\quad$ $(\alpha+\beta)\cancel{\Gamma(\alpha+\beta)}$

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

$\pi_0 \xrightarrow[n-s \text{ failures}]{s \text{ success}} \pi_1$

unif $(0,1)$ $\quad$ $Beta(s+1, n-s+1)$

$= Beta(1,1)$

$\hat{\theta}_B = \mathbb{E}_{\pi_1}\theta = \frac{s+1}{s+1+n-s+1}$

//

$P_{\pi_1}(\text{next obs success}) = \frac{s+1}{n+2}$

MLE estimate for θ

when you observe s success, $n-s$ failures.

$\hat{\theta}_{MLE} = \bar{x} = \frac{s}{n}$

Suppose n tosses of coin, you happen to see $s=0$ Heads.

$\hat{\theta}_{MLE} = \frac{0}{n} = 0$

$\hat{\theta}_B = \frac{0+1}{n+2} = \frac{1}{n+2} > 0$

- $\hat{\theta}_{MLE} = \frac{s}{n}$ was unbiased
  $= \bar{x}$ $\quad s = n\bar{x}$ $\quad \mathbb{E}\bar{x} = \theta$
- $\hat{\theta}_B = \frac{s+1}{n+2}$ is biased

$\mathbb{E}\hat{\theta}_B = \mathbb{E}\frac{n\bar{x}+1}{n+2} = \frac{n\mathbb{E}\bar{x}+1}{n+2} = \frac{\overset{(n+2)\theta-2\theta}{\overset{\parallel}{n\theta+1}}}{n+2} = \theta + \boxed{\frac{1-2\theta}{n+2}}$

$\neq \theta$ unless $\theta = \frac{1}{2}$.

But notice the bias $\frac{1-2\theta}{n+2} \to 0$ as $n \to \infty$

Consistency of Updates

$$\underset{\substack{\text{Unif}(0,1) \\ \hat{\theta}_0 = \mathbb{E}_\pi\theta}}{\pi_0} \xrightarrow{H} \underset{\substack{\text{Beta}(0,1) \\ \hat{\theta}_1}}{\pi_1} \xrightarrow{H} \pi_2 \longrightarrow \cdots \xrightarrow{H} \pi_s \xrightarrow{T} \pi_{s+1} \longrightarrow \cdots \xrightarrow{T} \pi_n$$

So far : Point estimate Ch6

Today : Interval estimate ch7
confidence Interval (CI)

$X_1, X_2, \dots, X_n \longrightarrow \boxed{\text{exp. obs.}} \longrightarrow x_1, x_2, \dots, x_n$

$n = \mathbb{E}X$

$\bar{x}$ sample mean

$[\bar{x}-\varepsilon, \bar{x}+\varepsilon]$ contains μ w/ some "level of confidence".

- ε also is a statistic, i.e. constructed from sample $x_1, \dots, x_n$.
- View CI $[\bar{x}-\varepsilon, \bar{x}+\varepsilon]$ as a stastic
- View CI $[\bar{X}-\varepsilon, \bar{X}+\varepsilon]$ as a RV.

  what's probability CI will contain μ?

  $\mathbb{P}(\mu \in [\bar{X}-\varepsilon, \bar{X}+\varepsilon])$ = Across many repetitional of the procedure, what function of CI's will contain μ? (Frequency)

  μ: not RV. $[\bar{X}-\varepsilon, \bar{X}+\varepsilon]$: random

$$\mathbb{P}(\bar{X}-\varepsilon \le \mu \le \bar{X}+\varepsilon)$$
$$= \mathbb{P}(\mu \le \bar{X}+\varepsilon,\ \bar{X}-\varepsilon \le \mu)$$
$$= \mathbb{P}(\mu-\varepsilon \le \bar{X},\ \bar{X} \le \mu+\varepsilon)$$
$$= \mathbb{P}(\mu-\varepsilon \le \bar{X} \le \mu+\varepsilon)$$
$$= \mathbb{P}(-\varepsilon \le \bar{X} \le \varepsilon)$$
$$= \mathbb{P}(|\bar{X}-\mu| \le \varepsilon)$$

Want to know : prob. dist. of RV $\bar{X}$

Example : $X_1, X_2, \dots, X_n$ i.i.d. $N(\mu, \sigma^2)$ σ^2 known, μ unknown

$[\bar{X}-\varepsilon, \bar{X}+\varepsilon]$ e.g. $\underbrace{\mathbb{P}(|\bar{X}-\mu| \le \varepsilon) = 95\%}_{95\%\ \text{CI},\ \alpha = 0.05} = 1-\alpha = 100(1-\alpha)\%$

$100(1-\alpha)\%$ CI

$$\bar{X} = \frac{1}{n}(X_1 + X_2 + \cdots + X_n) = N(\mathbb{E}\bar{X}, \operatorname{Var}(\bar{X})) = N\left(\mu, \frac{\sigma^2}{n}\right)$$

Fact : $X \sim N(\mu_1, \sigma_1^2)$, $Y \sim N(\mu_2, \sigma_2^2)$ independent

$$X+Y \sim N(\mu_1+\mu_2, \sigma_1^2+\sigma_2^2)$$

$$\mathbb{E}\bar{X} = \mu, \quad \operatorname{Var}(\bar{X}) = \frac{\sigma^2}{n}$$

$$\mathbb{P}(|\bar{X}-\mu| \le \varepsilon) = 1-\alpha$$

want to choose ε so that this is true.

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

$$Z = \frac{\bar{X}-\mu}{\sigma/\sqrt{n}} \sim N(\mathbb{E}Z, \operatorname{Var} Z) = N(0,1)$$

$$\mathbb{E}Z = 0, \quad \operatorname{Var}(Z) = 1$$

Derivation. $\mathbb{E}Z = \mathbb{E}\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\mathbb{E}\bar{X}-\mu}{\sigma/\sqrt{n}} = \frac{\mu-\mu}{\sigma/\sqrt{n}} = 0$

$$\operatorname{Var}(Z) = \mathbb{E}Z^2 - \cancelto{0}{(\mathbb{E}Z)^2} = \mathbb{E}Z^2 = \mathbb{E}\left(\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right)^2$$

$$= \frac{\mathbb{E}(\bar{X}-\mu)^2}{(\sigma/\sqrt{n})^2} = \frac{\operatorname{Var}(\bar{X})}{\sigma^2/n} = 1$$

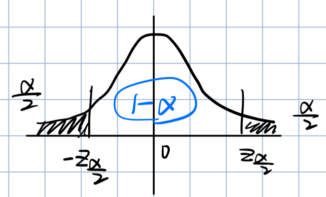
$$\mathbb{P}(|\bar{X}-\mu| \le \varepsilon)$$

$$= \mathbb{P}\left(\left|\frac{\bar{X}-\mu}{\sigma/\sqrt{n}}\right| \le \frac{\varepsilon}{\sigma/\sqrt{n}}\right)$$

$$= \mathbb{P}\left(|Z| \le \frac{\varepsilon}{\sigma/\sqrt{n}}\right) = 1-\alpha$$

$$= \mathbb{P}\left(-\frac{\varepsilon}{\sigma/\sqrt{n}} \le Z \le \underbrace{\frac{\varepsilon}{\sigma/\sqrt{n}}}_{z_{\frac{\alpha}{2}}}\right) = 1-\alpha$$

$$\varepsilon = z_{\frac{\alpha}{2}} \cdot \frac{\sigma}{\sqrt{n}}$$



$$CI = [\bar{X}-\varepsilon, \bar{X}+\varepsilon]$$

$$= [\bar{X} - z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}]$$

---

e.g. $\alpha = 0.05$, $1-\alpha = 0.95$, $z_{\frac{0.05}{2}} \approx 1.96$

$$[\bar{X} - 1.96\frac{\sigma}{\sqrt{n}}, \bar{X} + 1.96\frac{\sigma}{\sqrt{n}}]$$

More general: $X_1, X_2, \dots, X_n$ i.i.d. $\mathbb{E}X_1 = \mu$, $\operatorname{Var} X_1 = \sigma^2$

σ^2 known, μ unknown.

- But X_i's not necessarily normal
- How can we construct CI?

Do this approximate CI.

Use Central Limit Theorem! (P_{220} + P_{300})

Use of Large Numbers: in the setting before $\bar{X} = \frac{1}{n}(X_1 + \cdots + X_n) = \frac{1}{n}S_n$

then $\bar{X} \to \mathbb{E}X_1$ in probability.

Idea: $\operatorname{Var}(\bar{X}) = \frac{\sigma^2}{n} \to 0$ as $n \to \infty$

$$S_n = X_1 + \cdots + X_n$$

$$Z_n = \frac{S_n - \mathbb{E}S_n}{\sqrt{Var(S_n)}} \qquad \mathbb{E}Z_n = 0, \quad Var(Z_n) = 1$$

CLT says : $Z_n \to N(0,1)$ in distribution.

i.e. $F_{Z_n}(t) \to F_{N(0,1)}(t)$ as $n \to \infty$,

Take away : for large n, $Z_n \approx N(0,1)$

$$Z_n = \frac{S_n - \mathbb{E}S_n}{\sqrt{Var(S_n)}} = \frac{S_n - n\mu}{\sqrt{n\sigma^2}} = \frac{\overbrace{\frac{S_n}{n}}^{=\bar{X}} - \mu}{\sqrt{\frac{n\sigma^2}{n}}} = \frac{\bar{X} - \mu}{\sigma/\sqrt{n}} = \frac{\bar{X} - \mathbb{E}\bar{X}}{Var(\bar{X})} \approx N(0,1) \text{ for large } n.$$

$$\mathbb{E}S_n = n\mathbb{E}X_1 = n\mu$$

$$Var(S_n) = Var(X_1 + \cdots + X_n) = Var(X_1) + \cdots + Var(X_n)$$
$$= n\sigma^2$$

Use previous calculations,
an approximate CI of confidence level $(1-\alpha)100\%$ is $[\bar{X} - Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}, \bar{X} + Z_{\frac{\alpha}{2}}\frac{\sigma}{\sqrt{n}}]$

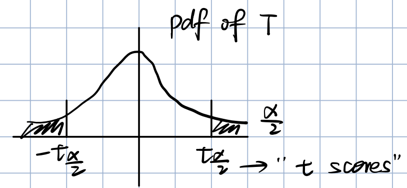
Unknown variance. $X_1, \dots, X_n$ i.i.d. mean μ unknown, variance σ^2 unknown.

- Use sample variance S^2 as approximation for σ^2

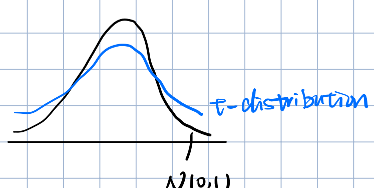
$$S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2 \qquad S = \sqrt{S^2} = \text{sample standard deviation}$$

Consider $\frac{\bar{X} - \mu}{S/\sqrt{n}} = T$

CI $[\bar{X} - \underbrace{t_{\frac{\alpha}{2}}}_{t_{\frac{\alpha}{2}}(n-1)}\frac{S}{\sqrt{n}}, \bar{X} + \underbrace{t_{\frac{\alpha}{2}}}_{t_{\frac{\alpha}{2}}(n-1)}\frac{S}{\sqrt{n}}]$ where



$T = \frac{\bar{X} - \mu}{S/\sqrt{n}}$ follows a t-distribution of $(n-1)$ degrees of freedom.



$$f_T(t) = \frac{1}{2}\left(1 + \frac{t^2}{n-1}\right)^{-\frac{n}{2}}$$

$$f_Z(t) = \frac{1}{\sqrt{2\pi}}e^{-\frac{t^2}{2}}$$

Updating belief using Bayesian Inference

- Hypothesis : unknown parameter θ
- Initial belief : prior distribution π_0 of θ
- Observation : $X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n$

  i.i.d. sample from pdf/pmf $f_X(\cdot \mid \theta)$
- Updated belief : posterior distribution π_1 of θ

$$\pi_1(\theta) := \pi_0(\theta \mid x_1, x_2, \cdots, x_n)$$

$$= \frac{\text{"}\mathbb{P}(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n \mid \theta)\; \pi_0(\theta)}{\text{"}\mathbb{P}(X_1 = x_1, X_2 = x_2, \cdots, X_n = x_n)\text{"}}$$

$$= \frac{1}{Z} \prod_{i=1}^{n} f_X(x_i \mid \theta)\; \pi_0(\theta)$$

where Z is a constant not depending on θ making this a pdf/pmf, i.e.

$$Z = \begin{cases} \sum \prod_{i=1}^{n} f_X(x_i \mid \theta)\; \pi_0(\theta), & \text{if } \theta \text{ discrete} \\ \int \prod_{i=1}^{n} f_X(x_i \mid \theta)\; \pi_0(\theta)\, d\theta, & \text{if } \theta \text{ continuous} \end{cases}$$

Z is called a "normalization constant"

Tip: pdf and normalization constant

- Given any function $g: \mathbb{R} \to [0,\infty)$ such that

  $$0 < \int g(x)\,dx < \infty$$

  we can turn it into a pdf by dividing it by

  $$Z = \int g(x)\,dx$$

  i.e. define

  $$f(x) = \frac{g(x)}{Z}$$

  then $\int f(x)\,dx = 1$.

- Conversely, if we know f is a pdf, and $f(x) = \frac{g(x)}{Z}$ for some constant Z, then

  $$Z = \int g(x)\,dx$$

  In particular, if $f_1(x) = \frac{g(x)}{Z_1}$ and $f_2(x) = \frac{g(x)}{Z_2}$ are both pdf's, then $Z_1 = Z_2$ and $f_1 = f_2$.

- This also applies to pmf, just change $\int$ to Σ.

- Consistency of updates

$$\Pi_0 \xrightarrow[x_1, x_2, \cdots, x_n]{\text{observe}} \Pi_1 \xrightarrow[x_{n+1}, \cdots, x_{n+m}]{\text{observe}} \Pi_2$$

$$\Pi_0 \xrightarrow[x_1, x_2, \cdots, x_n, x_{n+1}, \cdots, x_{n+m}]{\text{observe}} \Pi_2$$

Proof: $\Pi_1(\theta) = \frac{1}{Z_1} \prod_{i=1}^{n} f_X(x_i|\theta)\, \Pi_0(\theta)$

$$\Pi_2(\theta) = \frac{1}{Z_2} \prod_{i=n+1}^{n+m} f_X(x_i|\theta)\, \Pi_1(\theta)$$

$$= \frac{1}{Z_2} \prod_{i=n+1}^{n+m} f_X(x_i|\theta)\, \frac{1}{Z_1} \prod_{i=1}^{n} f_X(x_i|\theta)\, \Pi_0(\theta)$$

$$= \frac{1}{Z_1 Z_2} \prod_{i=1}^{n+m} f_X(x_i|\theta)\, \Pi_0(\theta)$$

Up to some constant factor, this is the same as the posterior distribution after observing $x_1, \cdots, x_n, \cdots, x_{n+m}$ given prior Π_0. But two pdf's are the same if they are the same up to a constant factor!

Bernoulli model and uniform prior

- Coin probability $p \in [0,1]$ of showing up Heads
- No information about p. So:

  prior $\pi_0 \sim$ uniform$(0,1)$

- Toss coin n times. Observe s Heads.
- What's the posterior π_1 of p?

Likelihood function $\mathbb{P}(x_1, x_2, \cdots, x_n \mid p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i}$

$$= p^s (1-p)^{n-s}$$

Bayes rule

$$\pi_1(p) := \pi_0(p \mid x_1, x_2, \cdots, x_n) = \frac{\mathbb{P}(x_1, x_2, \cdots, x_n \mid p)\, \pi_0(p)}{\mathbb{P}(x_1, x_2, \cdots, x_n)}$$

$$= \frac{1}{Z} p^s (1-p)^{n-s} \mathbb{1}_{p \in (0,1)}$$

- This looks like a <u>Beta</u> distribution:

pdf of Beta(α, β) $\quad (\alpha, \beta > 0)$

$$f_X(x) = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} x^{\alpha-1}(1-x)^{\beta-1} \mathbb{1}_{x \in (0,1)}$$

where $\Gamma(\alpha) := \int_0^\infty x^{\alpha-1} e^{-x} dx$, $\alpha > 0$

Match the parameters

$$\begin{aligned} s &= \alpha - 1 \\ n-s &= \beta - 1 \end{aligned} \longrightarrow \begin{aligned} \alpha &= s+1 \\ \beta &= n-s+1 \end{aligned}$$

pdf of $\text{Beta}(s+1, n-s+1)$

$$f(x) = \underbrace{\frac{\Gamma(n+2)}{\Gamma(s+1)\Gamma(n-s+1)}}_{\frac{1}{Z} =} x^s (1-x)^{n-s} \, 1_{x\in[0,1]}$$

constant, doesn't involve x

So $\pi_1 \sim \text{Beta}(s+1, n-s+1)$

- What if our prior is not uniform, but $\text{Beta}(\alpha, \beta)$? Posterior is $\text{Beta}(\alpha+s, \beta+n-s)$. (Ex 5 HW3)

- In the case α, β are integers, this can be seen by consistency of updates

$$\begin{array}{ccccc} \pi_0 & \xrightarrow[\beta-1 \text{ Tails}]{\alpha-1 \text{ Heads}} & \pi_1 & \overset{s \text{ Heads}}{\underset{n-s \text{ Tails}}{\dashrightarrow}} & \pi_2 \\ \text{Unif}(0,1) & & \text{Beta}(\alpha,\beta) & & \text{Beta}(\alpha+s, \beta+n-s) \\ \| & & & & \\ \text{Beta}(1,1) & & & & \end{array}$$

$\alpha+s-1$ Heads
$\beta+n-s-1$ Tails

Gamma function

- $\Gamma(\alpha+1) = \alpha\Gamma(\alpha)$

Use integration by parts,

$$\Gamma(\alpha+1) = \int_0^\infty x^\alpha e^{-x}\,dx$$

$$= \left[x^\alpha(-e^{-x})\right]_0^\infty - \int_0^\infty \alpha x^{\alpha-1}(-e^{-x})\,dx$$

$$= \alpha\int_0^\infty x^{\alpha-1}e^{-x}\,dx = \alpha\Gamma(\alpha)$$

- $\Gamma(1) = \int_0^\infty x^0 e^{-x}\,dx = 1$

$\Gamma(n+1) = n\Gamma(n) = \cdots = n!$

- Using this, we can compute

$$\frac{1}{Z} = \frac{\Gamma(n+2)}{\Gamma(s+1)\Gamma(n-s+1)} = \frac{(n+1)!}{s!\,(n-s)!} = (n+1)\cdot\frac{n!}{s!(n-s)!}$$

$$= (n+1)\binom{n}{s}$$

$$Z = \frac{1}{\binom{n}{s}}\,\frac{1}{n+1}.$$

Recall

$$Z = \mathbb{P}(x_1, x_2, \cdots, x_n)$$

Law of total probability $$= \int \mathbb{P}(x_1, x_2, \cdots, x_n \mid p)\, \pi_0(p)\, dp$$

$$= \int p^s (1-p)^{n-s}\, \mathbb{1}_{p \in [0,1]}\, dp$$

$$= \int_0^1 p^s (1-p)^{n-s}\, dp$$

Thus, $$\int_0^1 p^s (1-p)^{n-s}\, dp = \frac{1}{\binom{n}{s}} \frac{1}{n+1},$$

How could we show this in the first place?

$$\int_0^1 \binom{n}{s} p^s (1-p)^{n-s}\, dp = \frac{1}{n+1}$$

pmf of $\mathrm{Binom}(n,p)$!

Define $Y = X_1 + X_2 + \cdots + X_n$

Then conditioned on p, $X_1, \cdots, X_n$ i.i.d. $\mathrm{Ber}(p)$,

so $Y \sim \mathrm{Binom}(n,p)$,

$$\mathbb{P}(Y = s \mid p) = \binom{n}{s} p^s (1-p)^{n-s}$$

The L.H.S. computes $\mathbb{P}(Y=s)$ using the law of total probability.

- Thus, this equation is really saying that

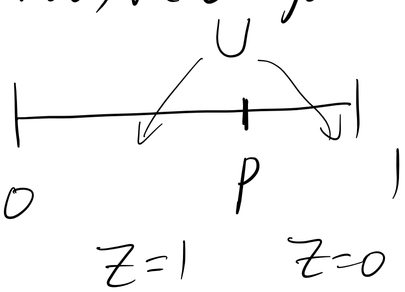
$$\underset{\text{unif}(0,1)}{p} \longrightarrow Y \sim \text{Binom}(n,p)$$

$$\mathbb{P}(Y=s) = \frac{1}{n+1}, \quad 0 \le s \le n.$$

i.e. $Y \sim \text{unif}(\{0,1,\dots,n\})$

- Here's a slick way to see this:

First, we can generate Ber(p) RV from unif(0,1) RV:



$$X = \mathbb{1}_{U \le p}$$

Now, let $U_1, U_2, \dots, U_n$ be i.i.d. unif(0,1) RV independent from p. (Remember p is also unif(0,1) RV)

$$X_1 = \mathbb{1}_{U_1 \le p}, \quad \dots, \quad X_n = \mathbb{1}_{U_n \le p},$$

are i.i.d. Ber(p) RV conditioned on p.

$$
\begin{aligned}
Y &= X_1 + X_2 + \cdots + X_n \\
&= \mathbb{1}_{U_1 \le p} + \mathbb{1}_{U_2 \le p} + \cdots + \mathbb{1}_{U_n \le p} \\
&= \#\{ i \in \{1, 2, \cdots, n\} : U_i \le p \} \\
&= \text{"rank" of } p \text{ among } \underbrace{p, U_1, U_2, \cdots, U_n}_{n+1 \text{ i.i.d. Ber}(p) \text{ RV's}}
\end{aligned}
$$

e.g. $Y = 2$ means

$$U^{(1)} < U^{(2)} < p < U^{(3)} < \cdots < U^{(n)}$$

Because the RV's are continuous, probability any two are equal is zero.

By symmetry, every "rank" is equally likely,
so Y is $\text{unif}(\{0, 1, \cdots, n\})$.

Bayesian Estimators

Recall:

$$\Pi_0 \xrightarrow[\text{s successes, } n-s \text{ failures}]{\text{observe}} \Pi_1$$

$\theta \sim Beta(\alpha, \beta)$ $\qquad\qquad$ $\theta | data \sim Beta(\alpha+s, \beta+n-s)$

- Want a point estimator $\hat{\theta}$ from the posterior, Π_1

  Consider the posterior mean

$$\hat{\theta} := \mathbb{E}_{\Pi_1} \theta = \mathbb{E}_{\Pi_0}[\theta \,|\, data\ x_1, x_2, \cdots, x_n]$$

- Let $X = X_{n+1}$ be a new independent observation. Then in fact:

$$\hat{\theta} = \mathbb{P}_{\Pi_1}(X=1) = \mathbb{P}_{\Pi_0}[X=1 \,|\, data\ x_1, x_2, \cdots, x_n]$$

Proof:
$$\begin{aligned}
\mathbb{P}_{\Pi_1}(X=1) &= \int \mathbb{P}_{\Pi_1}(X=1|\theta)\, \Pi_1(\theta)\, d\theta \\
&= \int \mathbb{P}(X=1|\theta)\, \Pi_1(\theta)\, d\theta \\
&= \int \theta\, \Pi_1(\theta)\, d\theta \\
&= \mathbb{E}_{\Pi_1} \theta .
\end{aligned}$$

- In the above,

$$\hat{\theta} = \text{mean of } Beta(\alpha+s, \beta+n-s) \text{ distribution}$$

- Mean of $Beta(\alpha, \beta)$ is

$$\mathbb{E}X = \frac{\alpha}{\alpha+\beta} \qquad\qquad \hat{\theta} = \frac{\alpha+s}{\alpha+\beta+n}$$

• Calculation:

$$\mathbb{E}X = \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x \cdot x^{\alpha-1} (1-x)^{\beta-1} dx$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \int_0^1 x^{(\alpha+1)-1} (1-x)^{\beta-1} dx$$

$$= \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\cancel{\Gamma(\beta)}} \frac{\Gamma(\alpha+1)\cancel{\Gamma(\beta)}}{\Gamma(\alpha+\beta+1)}$$

$$= \frac{\cancel{\Gamma(\alpha+\beta)}}{\cancel{\Gamma(\alpha)}} \frac{\alpha \cancel{\Gamma(\alpha)}}{(\alpha+\beta)\cancel{\Gamma(\alpha+\beta)}} = \frac{\alpha}{\alpha+\beta}$$

• Recall: $\text{Beta}(1,1) = \text{unif}(0,1)$

$$\Pi_0 \xrightarrow[\text{s successes, n-s failures}]{\text{observe}} \Pi_1$$

$\theta \sim \text{unif}(0,1)$

$\theta | \text{data} \sim \text{Beta}(s+1, n-s+1)$

$$\hat{\theta} = \mathbb{E}_{\Pi_1}(\theta) = \frac{s+1}{n+2}$$

This is known as **Laplace Rule of Succession**

• This might seem paradoxical at first:

Why not $\hat{\theta} = \frac{s}{n} = \bar{x}$ instead? This was the MLE estimator

e.g. If toss coin 10 times, see 6 Heads,

why not $\hat{\theta} = \frac{6}{10}$, but $\hat{\theta} = \frac{7}{12}$?

- Compare the two estimators

$$\hat{\theta}_{MLE} = \frac{S}{n}, \qquad \hat{\theta}_{Bayes} = \frac{S+1}{n+2}$$

If $S=0$, i.e. toss coin n times and gets no Heads

$$\hat{\theta}_{MLE} = \frac{0}{n} = 0, \qquad \hat{\theta}_{Bayes} = \frac{1}{n+2}$$

In a sense, $\hat{\theta}_{MLE}$ is "over confident" about the value of θ.
If $\theta > 0$ but $\theta < \frac{1}{n}$, then it's quite possible to see no Heads, but it's extreme to estimate θ as 0.

- **Laplace Rule of Succession** says we should imagine there was originally a Head and a Tail before we start, so "# Heads" $= S+1$, "# Tosses" $= n+2$, and use

$$\hat{\theta} = \frac{S+1}{n+2} \text{ to estimate } \theta.$$

This is equivalent to Bayesian update from uniform prior and then take the posterior mean.

- We know $\hat{\theta}_{MLE} = \frac{S}{n} = \bar{x}$ is unbiased.

But $\hat{\theta}_{Bayes} = \frac{S+1}{n+2} = \frac{n\bar{x}+1}{n+2}$ is biased:

$$\mathbb{E}_\theta \frac{n\bar{x}+1}{n+2} = \frac{n\,\mathbb{E}\bar{x}+1}{n+2} = \frac{n\theta+1}{n+2} = \theta + \frac{1-2\theta}{n+2} \neq \theta$$

unless $\theta = \frac{1}{2}$.

$$\pi_0 \xrightarrow{H} \pi_1 \xrightarrow{H} \pi_2 \longrightarrow \cdots \xrightarrow{H} \pi_s$$

$\mathrm{Unif}(0,1) = \mathrm{Beta}(1,1)$, $\mathrm{Beta}(2,1)$, $\mathrm{Beta}(3,1)$, $\mathrm{Beta}(s+1,1)$

$$\xrightarrow{T} \pi_{s+1} \xrightarrow{T} \pi_{s+2} \longrightarrow \cdots \xrightarrow{T} \pi_n$$

$\mathrm{Beta}(s+1,2)$, $\mathrm{Beta}(s+1,3)$, $\mathrm{Beta}(s+1, n-s+1)$

$$\hat\theta_0 = \frac{1}{2},\quad \hat\theta_1 = \frac{1+1}{1+2} = \frac{2}{3},\quad \hat\theta_2 = \frac{2+1}{2+2} = \frac{3}{4},\ \cdots,\ \hat\theta_s = \frac{s+1}{s+2}$$

$$\hat\theta_{s+1} = \frac{s+1}{s+3},\quad \hat\theta_{s+2} = \frac{s+1}{s+4},\ \cdots,\ \hat\theta_n = \frac{s+1}{n+2}$$

$$\begin{aligned}
&\mathbb{P}(X_1=H, X_2=H, \cdots, X_s=H, X_{s+1}=T, \cdots, X_n=T) \\
&= \mathbb{P}(X_1=H)\,\mathbb{P}(X_2=H \mid X_1=H) \cdots \mathbb{P}(X_s=H \mid X_1, \cdots, X_{s-1}=H) \\
&\quad \mathbb{P}(X_{s+1}=T \mid X_1, \cdots, X_s=H) \cdots \mathbb{P}(X_n=T \mid X_1, \cdots, X_s=H, X_{s+1}, \cdots, X_{n-1}=T) \\
&= \mathbb{P}_{\pi_0}(X_1=H)\,\mathbb{P}_{\pi_1}(X_2=H) \cdots \mathbb{P}_{\pi_{n-1}}(X_n=T) \\
&= \hat\theta_0\,\hat\theta_1 \cdots \hat\theta_{s-1}(1-\hat\theta_s) \cdots (1-\hat\theta_n) \\
&= \frac{1}{2} \times \frac{2}{3} \times \cdots \times \frac{s}{s+1} \times \left(1-\frac{s+1}{s+2}\right) \times \cdots \times \left(1-\frac{s+1}{n+1}\right) \\
&\qquad \frac{1}{s+2} \times \cdots \times \frac{n-s}{n+1} \\
&= \frac{s!\,(n-s)!}{(n+1)!} \\
&= \frac{1}{(n+1)\binom{n}{s}}
\end{aligned}$$