1.1 Classical and frequentist probability

$P(X=4)$ $\quad$ $P(\text{drop packet})$

$P(\text{die} = \text{fair})$ $\quad$ $P(Y_1 > Y_2)$

$P(\text{rain})$

Classical – equally likely

$P(X=4) = 1/6$ for a die

$$P(X_1+X_2=4) = \frac{3}{36} = \frac{1}{12} = \frac{\#\text{ possible}}{\#\text{ equally likely outcomes}}$$

Frequentist – relative frequency in infinite sequence

$P(X=4) = 1/6$

$P(\text{drop packets}) = \frac{1}{10{,}000}$ (lose 1 in 10,000)

$P(\text{rain})$ requires an infinite sequence

$P(\text{fair}) = \{0,1\}$ (fair or not regardless of all of the rolls)

Bayesian – personal perspective

1.2 Bayesian probability and coherence

(1) If rain, win \$4, If no rain, lose \$1 $\quad$ odds 4:1

(2) If rain lose \$4, If no rain win \$1

If this is a fair game: $P(\text{rain}) = \frac{1}{1+4} = \frac{1}{5}$

Expected return: $4\left(\frac{1}{5}\right) - 1\left(\frac{4}{5}\right) = 0$ $\quad$ (1)

$1\left(\frac{4}{5}\right) - 4\left(\frac{1}{5}\right) = 0$ $\quad$ (2)

Coherence: Must follow all rules of probability

2.1 Conditional probability

$$P(A|B) = \frac{P(A\cap B)}{P(B)}$$

30 students
9 females
12 computer science of which 4 female

$$P(F) = \frac{9}{30} = \frac{3}{10} \qquad P(CS) = \frac{12}{30} = \frac{2}{5}$$

$$P(F\cap CS) = \frac{4}{30} = \frac{2}{15}$$

$$P(F|CS) = \frac{P(F\cap CS)}{P(CS)} = \frac{\frac{2}{15}}{\frac{2}{5}} = \frac{1}{3}$$

$$P(F|CS^c) = \frac{P(F\cap CS^c)}{P(CS^c)} = \frac{5/30}{18/30} = 5/18$$

Independence

$$P(A|B) = P(A) \Rightarrow P(A\cap B) = P(A)P(B)$$

$$P(F|CS) \neq P(F) \Rightarrow \text{not independent}$$

2.2 Bayes' Theorem

$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)} = \frac{P(A\cap B)}{P(B)}$$

$$P(CS|F) = \frac{P(F|CS)\,P(CS)}{P(F|CS)\,P(CS) + P(F|CS^c)\,P(CS^c)}$$

$$= \frac{\left(\frac{1}{3}\right)\left(\frac{2}{5}\right)}{\left(\frac{1}{3}\right)\left(\frac{2}{5}\right) + \left(\frac{5}{18}\right)\left(\frac{3}{5}\right)} = \frac{4}{9}$$

$$P(CS|F) = \frac{P(CS\cap F)}{P(F)} = \frac{4/30}{9/30} = \frac{4}{9}$$

$$P(+|HIV) = .977 \qquad P(HIV) = .0026$$

$$P(-|\text{no } HIV) = .926$$

$$P(HIV|+) = \frac{P(+|HIV)\,P(HIV)}{P(+|HIV)P(HIV) + P(+|\text{no } HIV)\,P(\text{no } HIV)}$$

$$= \frac{(.977)(.0026)}{(.977)(.0026) + (1-.926)(1-.0026)} = .033$$

3.1 Bernoulli and binomial distributions

Bernoulli

$X \sim B(p) \qquad P(X=1) = p$
$\qquad\qquad\quad\; P(X=0) = 1-p$

$$f(X=x|p) = f(x|p) = p^x(1-p)^{1-x} I_{\{x \in \{0,1\}\}}(x)$$

Expected value

$$E[X] = \sum_x x P(X=x) = (1)p + (0)(1-p) = p$$

$$\mathrm{Var}(X) = p(1-p)$$

Binomial

$$X \sim \mathrm{Bin}(n,p)$$

$$P(X=x|p) = f(x|p) = \binom{n}{x} p^x (1-p)^{n-x}$$

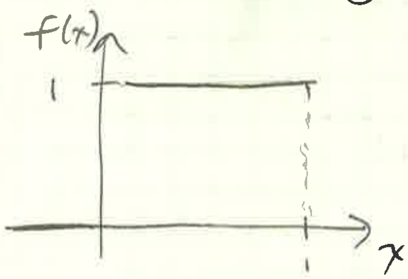
$$\binom{n}{x} = \frac{n!}{x!(n-x)!} \quad \text{for } x \in \{0,1,\ldots,n\}$$

$$E[X] = np \quad , \quad \mathrm{Var}(X) = np(1-p)$$

3.2 Uniform distribution

$X \sim U[0,1]$

$$f(x) = \begin{cases} 1 & \text{if } x \in [0,1] \\ 0 & \text{otherwise} \end{cases} = I_{\{0 \le x \le 1\}}(x)$$



$$P(0 < X < \tfrac{1}{2}) = \int_0^{1/2} f(x)dx = \int_0^{1/2} dx = \frac{1}{2}$$

$$P(X = \tfrac{1}{2}) = 0$$

$$\int_{-\infty}^{\infty} f(x)dx = 1 \quad , \quad f(x) \ge 0$$

$$E[X] = \int_{-\infty}^{\infty} x f(x) dx$$

$$E[g(X)] = \int_{-\infty}^{\infty} g(x) f(x) dx$$

$$E[cX] = cE(X) \quad , \quad E[X+Y] = E[X] + E[Y]$$

$$\text{if } X \perp Y \Rightarrow E[XY] = E[X]E[Y]$$

3.3 Exponential and normal distributions

Exponential

$X \sim Exp(\lambda)$

$f(x|\lambda) = \lambda e^{-\lambda x}$ for $x \geq 0$

$E[X] = \frac{1}{\lambda}$, $Var(X) = \frac{1}{\lambda^2}$

Uniform $X \sim U[\theta_1, \theta_2]$

$f(x|\theta_1, \theta_2) = \frac{1}{\theta_2 - \theta_1} I_{\{\theta_1 \leq x \leq \theta_2\}}$

Normal

$X \sim N(\mu, \sigma^2)$

$f(x|\mu, \sigma^2) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left(\frac{-1}{2\sigma^2}(x-\mu)^2 \right)$

$E[X] = \mu$, $Var(X) = \sigma^2$

4.1 Confidence Intervals

There is a fixed value of p

Flip a coin 44H, 56T. Is this a fair coin?

$X_i \sim B(p)$

By CLT $\sum_{i=1}^{100} X_i \overset{\cdot}{\sim} N(100p, 100p(1-p))$

$100p - 1.96\sqrt{100p(1-p)}$ and $100p + 1.96\sqrt{100p(1-p)}$

Observe $\sum x_i = 44 \Rightarrow \hat{p} = \frac{44}{100} = 0.44$

CI $44 \pm 1.96\sqrt{44(.56)} = 44 \pm 9.7 = [34.3, 53.7)$

$\Leftrightarrow$ 95% confident $p \in (.343, .537)$

If we repeat this experiment an arbitrary number of times, then, on average 95% of the intervals that we make will contain the true p. Thus, the frequentist probability $P(p \in CI) = \{0,1\}$ since p is a fixed value

4.2 Likelihood function and maximum likelihood

$Y_i \overset{iid}{\sim} B(\theta)$, $P(Y_i = 1) = \theta$

$$
\begin{aligned}
P(\underset{\sim}{Y} = \underset{\sim}{y} \mid \theta) &= P(Y_1 = y_1, Y_2 = y_2, \ldots, Y_n = y_n \mid \theta) \\
&= P(Y_1 = y_1 \mid \theta) \cdots P(Y_n = y_n \mid \theta) \\
&= \prod_{i=1}^{n} P(Y_i = y_i \mid \theta) = \prod_{i=1}^{n} \theta^{y_i} (1-\theta)^{1-y_i}
\end{aligned}
$$

Likelihood: $L(\theta \mid \underset{\sim}{y}) = \prod_{i=1}^{n} \theta^{y_i} (1-\theta)^{1-y_i}$

MLE: $\hat{\theta} = \operatorname{argmax} L(\theta \mid \underset{\sim}{y})$

$$
\begin{aligned}
\ell(\theta) = \log L(\theta \mid \underset{\sim}{y}) &= \log \left[\prod_{i=1}^{n} \theta^{y_i} (1-\theta)^{1-y_i}\right] \\
&= \sum_{i=1}^{n} \log\left[\theta^{y_i} (1-\theta)^{1-y_i}\right] = \sum_{i=1}^{n} \left[y_i \log\theta + (1-y_i)\log(1-\theta)\right] \\
&= \left(\sum_{i=1}^{n} y_i\right) \log\theta + \left(\sum_{i=1}^{n} (1-y_i)\right) \log(1-\theta)
\end{aligned}
$$

4.3 Computing the MLE

$$\ell'(\theta) = \frac{1}{\theta}\sum y_i - \frac{1}{1-\theta}\sum(1-y_i) \stackrel{\text{set}}{=} 0$$

$$\Rightarrow \frac{\sum y_i}{\hat\theta} = \frac{\sum(1-y_i)}{1-\hat\theta}$$

$$\Rightarrow \hat\theta = \frac{1}{n}\sum y_i = \hat p = \frac{72}{400} = .18$$

Approx CI: $\hat\theta \pm 1.96\sqrt{\frac{\hat\theta(1-\hat\theta)}{n}}$

$\hat\theta \,\dot\sim\, N\left(\theta, \frac{1}{I[\hat\theta]}\right)$ $\quad$ $I(\theta)$ is the Fisher Information

$I(\theta) = \frac{1}{\theta(1-\theta)}$ for a Bernoulli

4.4 Computing the MLE: Examples

$X_i \stackrel{iid}{\sim} Exp(\lambda)$

$$f(\underset{\sim}{x}|\lambda) = \prod_{i=1}^{n} \lambda e^{-\lambda x_i} = \lambda^n e^{-\lambda\sum x_i}$$

$$L(\lambda|\underset{\sim}{x}) = \lambda^n e^{-\lambda\sum x_i}$$

$$\ell(\lambda) = n\log\lambda - \lambda\sum x_i$$

$$\ell'(\lambda) = \frac{n}{\lambda} - \sum x_i \stackrel{\text{set}}{=} 0 \quad \Rightarrow \hat\lambda = \frac{n}{\sum x_i} = \frac{1}{\bar{x}}$$

$X_i \stackrel{iid}{\sim} U[0,\theta]$

$$f(\underset{\sim}{x}|\theta) = \prod_{i=1}^{n} \frac{1}{\theta} I_{\{0\le x_i\le\theta\}}$$

$$L(\theta|\underset{\sim}{x}) = \theta^{-n} I_{\{0\le \min x_i \le \max x_i \le \theta\}}$$

$$L'(\theta) = -n\theta^{-(n+1)} I_{\{0\le \min x_i \le \max x_i \le \theta\}}$$

$$\hat\theta = \max x_i$$

5.1 Inference Example: Frequentist

$\theta = \{\text{fair}, \text{loaded}\}$

$X \sim Bin(5, ?)$ 5 flips, what is p?

$$f(x|\theta) = \begin{cases} \binom{5}{x}\left(\frac{1}{2}\right)^5 & \text{if } \theta = \text{fair} \\ \binom{5}{x}(.7)^x(.3)^{5-x} & \text{if } \theta = \text{loaded} \end{cases}$$

$$= \binom{5}{x}(.5)^5 I_{\{\theta = \text{fair}\}} + \binom{5}{x}(.7)^x(.3)^{5-x} I_{\{\theta = \text{loaded}\}}$$

let $X = 2$

$$f(\theta | X=2) = \begin{cases} 0.3125 & \text{if } \theta = \text{fair} \\ 0.1323 & \text{if } \theta = \text{loaded} \end{cases}$$

MLE $\hat{\theta} = \text{fair}$

$P(\theta = \text{fair} | X=2) = P(\theta = \text{fair}) = \{0, 1\}$ since it is a fixed coin

5.2 Inference Example: Bayesian

Prior $P(\theta = \text{loaded}) = 0.6$

$$f(\theta|x) = \frac{f(x|\theta) f(\theta)}{\sum_\theta f(x|\theta) f(\theta)}$$

$$= \frac{\binom{5}{x}\left[\left(\frac{1}{2}\right)^5(.4) I_{\{\theta=\text{fair}\}} + (.7)^x(.3)^{5-x}(.6) I_{\{\theta=\text{loaded}\}}\right]}{\sum_\theta \binom{5}{x}\left[\left(\frac{1}{2}\right)^5(.4) I_{\{\theta=\text{fair}\}} + (.7)^x(.3)^{5-x}(.6) I_{\{\theta=\text{loaded}\}}\right]}$$

$$= \frac{0.0125 I_{\{\theta=\text{fair}\}} + 0.0079 I_{\{\theta=\text{loaded}\}}}{0.0125 + 0.0079}$$

$$= 0.612 I_{\{\theta=\text{fair}\}} + 0.388 I_{\{\theta=\text{loaded}\}}$$

$P(\theta = \text{loaded} | X=2) = .388$

$P(\theta = \text{loaded}) = \frac{1}{2} \Rightarrow P(\theta = \text{loaded} | X=2) = .297$ (different priors)

$P(\theta = \text{loaded}) = 0.9 \Rightarrow P(\theta = \text{loaded} | X=2) = .792$

5.3 Continuous version of Bayes' Theorem

$$f(\theta|y) = \frac{f(y|\theta)f(\theta)}{f(y)} = \frac{f(y|\theta)\,f(\theta)}{\int f(y|\theta)\,f(\theta)\,d\theta} = \frac{\text{likelihood} \times \text{prior}}{\text{normalizing constant}}$$

$$\propto \text{likelihood} \times \text{prior}$$

$$\theta \sim U[0,1] \quad , \; f(\theta) = I_{\{0 \le \theta \le 1\}}$$

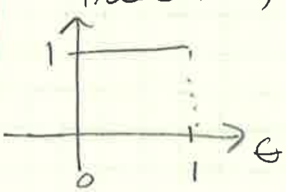
$$f(\theta|Y=1) = \frac{\theta^1(1-\theta)^0 I_{\{0\le\theta\le 1\}}}{\int_{-\infty}^{\infty} \theta^1(1-\theta)^0 I_{\{0\le\theta\le 1\}}} = \frac{\theta\, I_{\{0\le\theta\le 1\}}}{\int_0^1 \theta \, d\theta} = 2\theta\, I_{\{0\le\theta\le 1\}}$$

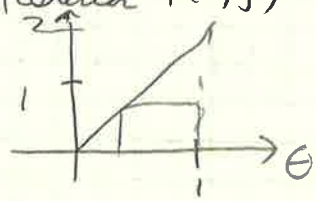
$$f(\theta|y) \propto f(y|\theta)f(\theta) \propto \theta\, I_{\{0\le\theta\le 1\}}$$

$$\Rightarrow f(\theta|Y=1) = 2\theta\, I_{\{0\le\theta\le 1\}}$$

5.4 Posterior Intervals

Prior $f(\theta)$ Posterior $f(\theta,y)$





Prior interval estimates

$$P(0.025 < \theta < 0.975) = .95 \quad , \quad P(\theta > 0.05) = 0.95$$

Posterior interval estimates

$$P(0.025 < \theta < 0.975) = \int_{0.025}^{0.975} 2\theta\, d\theta = .975^2 - .025^2 = .95$$

$$P(\theta > 0.05) = 1 - 0.05^2 = .9975$$

Equal tailed

$$P(\theta < q \mid Y=1) = \int_0^q 2\theta\, d\theta = q^2$$

$$P(\sqrt{0.025} < \theta < \sqrt{.975}) = P(0.158 < \theta < 0.987) = 0.95$$

HPD $\quad P(\theta > \sqrt{.05} \mid Y=1) = P(\theta > .224 \mid Y=1) = 0.95$