Lecture 2: Priors and Inversion Sampling

sem 2, 2018

Outline

1. Different types of priors
    - Conjugate priors, improper priors, Jeffreys' priors
2. Monte Carlo Methods
    - Inversion sampling.

Summary: Bayesian updating

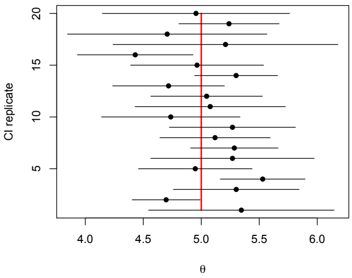
There are 4 key steps in the Bayesian approach:

- Specification of a likelihood model $L(x|\theta)$;
- Determiniation/elicitation of a suitable prior distribution $\pi(\theta)$;
- Calculation of the posterior distribution via Bayes' Theorem

$$\pi(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\int_{\Theta} L(x|\theta)\pi(\theta)\mathrm{d}\theta}$$

- Draw inference from the posterior.

Confidence/Credible Intervals



Classical Interpretation:

"In the long run over many experiments, 95% of CI's will contain the true parameter"

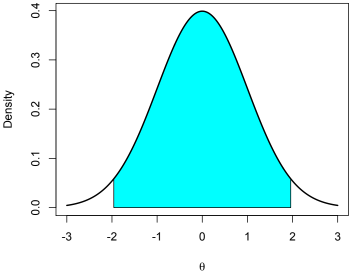
(But your single CI either will or won't)

Bayesian Interpretation:

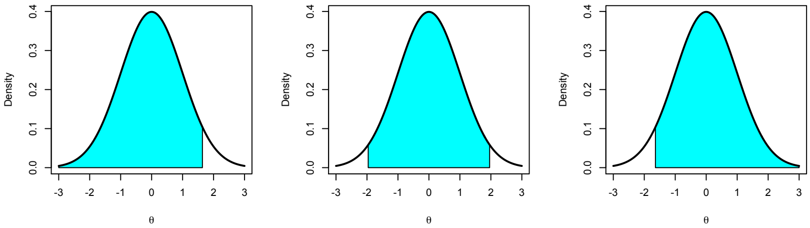
"There is a 95% probability that θ is in your (single) credible interval"

Very intuitive and simple.

Credible versus Confidence Interval



A comment on Credible Intervals



Credible intervals are not unique – there are an ∞ number!

All have exactly the same interpretation (the choice is yours).

High Density Region (HDR) intervals:

- The credible interval with the shortest width
- Not always the central interval e.g. if posterior is skewed

High Density Region interval

- Suppose we consider the Weibull pdf $\alpha x^{\alpha-1}\exp(-x^{\alpha}), x > 0$ with $\alpha = 4$ and seek a 95% HDR.
- The cdf is $F(x) = 1 - \exp(-x^{\alpha})$. We need to find the smallest width $b - a$ of the interval $[a, b]$ such that

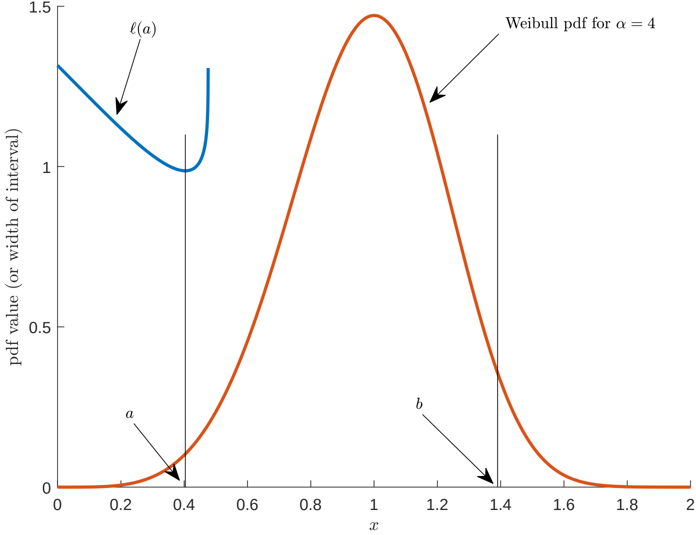
$$F(b) - F(a) = \exp(-a^{\alpha}) - \exp(-b^{\alpha}) = 0.95 \ .$$

- Eliminating $b = (-\ln(\exp(-a^{\alpha}) - 0.95))^{1/\alpha}$, we wish to find $a < (-\ln(0.95))^{1/\alpha}$ that minimizes the length:

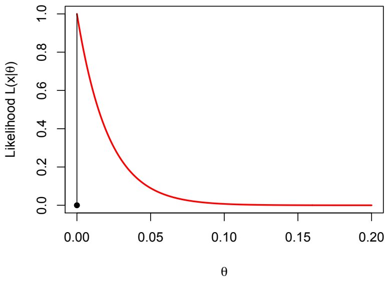
$$\ell(a) = (-\ln(\exp(-a^{\alpha}) - 0.95))^{1/\alpha} - a$$

- We can find the minimum of ℓ graphically or using calculus: $a \approx 0.4036$ and $\ell(0.4036) \approx 0.9868$.

High Density Region



Non-regular likelihoods



Observed data: $r = 0$, $n = 47$

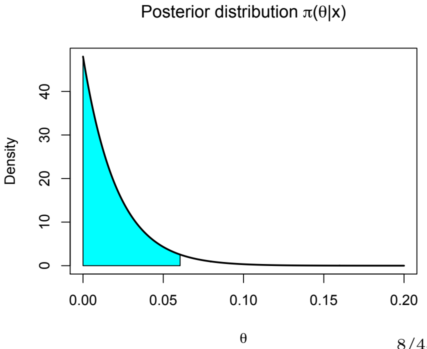
$$r \sim \mathrm{Bin}(n, \theta)$$

The likelihood:

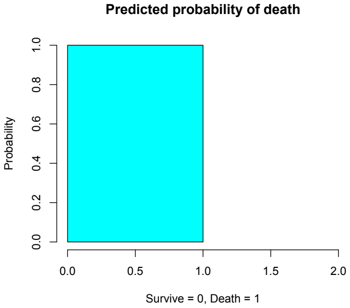
- $\hat{\theta} = 0$ is still well defined
- But $\hat{\theta} \sim \mathsf{N}(\theta, \sigma_0^2)$ fails
- How to get a CI for θ?

The Bayesian approach

- $\theta \sim \pi(\theta) = \mathsf{U}(0, 1)$ prior
- No requirement of "standard" likelihood asymptotics.
- The posterior contains all required information.



Predictive distributions



Hospital Example:

Classical predictive distribution will produce:

- 100% prediction for survival
- 0% prediction for death

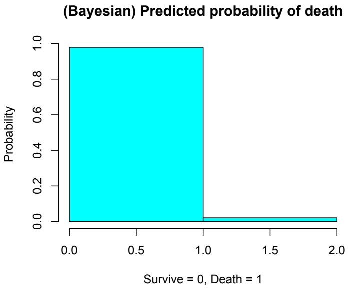
This is completely unrealistic

Bayesian predictive procedure:

- $\theta \sim \pi(\theta \mid x)$ draw from posterior
- $Y \sim \mathsf{Ber}(\theta)$

At least some predicted possibility of future deaths

Accounts for parameter uncertainty.



Outline

1. Recap
2. Different types of priors
   - Conjugate priors, improper priors, Jeffreys' priors
3. Monte Carlo Methods
   - Inversion sampling.

Effect of prior on posterior computation

In order to compute probabilities based on the posterior

$$\pi(\theta|x) = \frac{L(x|\theta)\pi(\theta)}{\int_\Theta L(x|\theta)\pi(\theta)\mathrm{d}\theta}$$

we need to be able to evaluate the normalising constant.

Example:
Suppose $x_1, \ldots, x_n \sim \text{Poisson}(\theta)$ and that we believe $0 < \theta \leq 1$ with all values equally likely. Hence $\pi(\theta) = 1, 0 < \theta \leq 1.$
Here the normalising constant is proportional to:

$$\int_0^1 \exp\left(-n\theta\right) \theta^{\sum_{i=1}^n x_i} \mathrm{d}\theta$$

which can only be evaluated numerically.

The point:

Even simple priors can lead to awkward numerical problems.

Conjugate priors

Conjugate prior definition:

If

- $\mathcal{F}$ is a class of sampling distributions (likelihoods) $L(x \,|\, \theta)$, and
- $\mathcal{P}$ is a class of prior distributions $\pi(\theta)$ for θ,

then the class $\mathcal{P}$ is conjugate for $\mathcal{F}$ if

$$\pi(\theta \,|\, x) \in \mathcal{P} \qquad \text{for all } L(\cdot \,|\, \theta) \in \mathcal{F} \text{ and } \pi(\cdot) \in \mathcal{P}.$$

i.e. Conjugate priors permit the posterior distribution to stay in the same family as the prior.

Viable if family $\mathcal{P}$ is flexible enough to describe prior beliefs.

Conjugate priors

Example #1: Binomial model

Suppose we have $X \sim \text{Binomial}(n, \theta)$ and we wish to make inference on $0 \leq \theta \leq 1$.

Here

$$L(x|\theta) = \left(\begin{array}{c} n \\ x \end{array} \right) \theta^x (1-\theta)^{n-x}, \quad x = 0, \ldots, n.$$

While $\pi(\theta)$ will vary from problem to problem, we want to choose a prior family that will simplify posterior computations.

Specify $\theta \sim \text{Beta}(a, b)$ so that

$$\begin{aligned} \pi(\theta) &= \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} \theta^{a-1} (1-\theta)^{b-1} \qquad 0 \leq \theta \leq 1 \\ &\propto \theta^{a-1} (1-\theta)^{b-1}. \end{aligned}$$

Conjugate priors

Example #1: Binomial model (cont.)

Then, by Bayes' Theorem:

$$\begin{aligned}\pi(\theta \mid x) &\propto L(x \mid \theta)\pi(\theta) \\ &\propto \theta^{x}(1-\theta)^{n-x} \times \theta^{a-1}(1-\theta)^{b-1} \\ &= \theta^{a+x-1}(1-\theta)^{b+n-x-1}.\end{aligned}$$

Since we know $\pi(\theta \mid x)$ is a proper density function ($\int \pi(\theta \mid x)\mathrm{d}\theta = 1$), then we must have

$$\theta \mid x \sim \text{Beta}(a+x, b+n-x).$$

- Effect of data: modify Beta parameters from (a, b) to $(a+x, b+n-x)$;
- Staying within Beta family means posterior computation is easy.

Conjugate priors

Example #2: Poisson model

Suppose $X_1, \ldots, X_n$ are *iid* $\text{Poi}(\theta)$ variables.
Then

$$
\begin{aligned}
L(x \mid \theta) &= \prod_{i=1}^{n} \exp(-\theta)\theta^{x_i}/x_i! \\
&\propto \exp(-n\theta)\theta^{\sum_{i=1}^{n} x_i}.
\end{aligned}
$$

Suppose prior beliefs can be represented by $\theta \sim \text{Gamma}(a, b)$, so that

$$
\begin{aligned}
\pi(\theta) &= \frac{a^b}{\Gamma(a)}\theta^{a-1}\exp(-b\theta) \qquad \theta > 0 \\
&\propto \theta^{a-1}\exp(-b\theta).
\end{aligned}
$$

Conjugate priors

Example #2: Poisson model (cont.)

Then by Bayes' Theorem:

$$\begin{aligned}
\pi(\theta \,|\, x) &\propto L(x \,|\, \theta)\pi(\theta) \\
&\propto \exp(-n\theta)\theta^{\sum_i x_i}\theta^{a-1}\exp(-b\theta) \\
&= \theta^{a+\sum_i x_i - 1}\exp(-(b+n)\theta).
\end{aligned}$$

Hence, we must have

$$\theta \,|\, x \sim \text{Gamma}(a + n\bar{x}_n, b + n).$$

- Effect of data: modify Gamma parameters from (a, b) to $(a + n\bar{x}_n, b + n)$;
- Stays within Gamma family.

Conjugate priors

Example #3: Gamma model

Suppose $X_1, \ldots, X_n$ are *iid* Gamma(k, θ) variables, with k known/fixed.
Then

$$\begin{aligned} L(x \mid \theta) &= \prod_{i=1}^{n} \frac{\theta^k}{\Gamma(k)} {x_i}^{k-1} \exp(-x_i\theta) \\ &\propto \theta^{kn} \exp(-\theta n\bar{x}) \end{aligned}$$

Suppose prior beliefs can again be represented by $\theta \sim \text{Gamma}(a, b)$:

$$\pi(\theta) = \theta^{a-1} \exp(-b\theta).$$

Then by Bayes' Theorem:

$$\begin{aligned} \pi(\theta \mid x) &\propto \theta^{kn} \exp(-\theta n\bar{x})\theta^{a-1} \exp(-b\theta) \\ &= \theta^{a+kn-1} \exp(-(b + n\bar{x})\theta) \\ &\sim \text{Gamma}(a + kn, b + n\bar{x}). \end{aligned}$$

Conjugate priors

When can conjugate priors be found?

A: The only easy case is for models in the exponential family:

$$f(x \mid \theta) = h(x)g(\theta)\exp\{t(x)c(\theta)\}$$

for functions h, g, t and c such that $\int f(x \mid \theta)\mathrm{d}x = 1$.

Includes: Exponential, Poisson, one-parameter Gamma, Binomial, Normal (with known variance).

Conjugate priors

In this case, with prior $\pi(\theta)$, we have:

$$\begin{aligned}
\pi(\theta \,|\, x) &\propto L(x \,|\, \theta)\pi(\theta) \\
&= \pi(\theta) \prod_{i=1}^{n} h(x_i) g(\theta) \exp\{t(x_i) c(\theta)\} \\
&\propto \pi(\theta) g(\theta)^n \exp\left\{c(\theta) \textstyle\sum_i t(x_i)\right\}.
\end{aligned}$$

Which means that if we choose

$$\pi(\theta) \propto g(\theta)^d \exp\{bc(\theta)\},$$

for some d and b, then we obtain

$$\begin{aligned}
\pi(\theta \,|\, x) &\propto g(\theta)^{d+n} \exp\left\{c(\theta)[\textstyle\sum_i t(x_i) + b]\right\} \\
&= g(\theta)^{\tilde{d}} \exp\left\{\tilde{b}\, c(\theta)\right\},
\end{aligned}$$

which is in the exponential family (with modified parameters).

Conjugate priors

Example : Binomial model

Here we have

$$
\begin{aligned}
f(x \mid \theta) &= \binom{n}{x} \theta^x (1-\theta)^{n-x}, \\
&= \binom{n}{x} (1-\theta)^n \exp\left\{ x \log\left(\frac{\theta}{1-\theta} \right) \right\}.
\end{aligned}
$$

So here we have: $h(x) = \binom{n}{x}$, $g(\theta) = (1-\theta)^n$, $t(x) = x$ and $c(\theta) = \log\left(\frac{\theta}{1-\theta}\right)$.
Hence we need to construct a prior of the form

$$
\begin{aligned}
\pi(\theta) &\propto [(1-\theta)^n]^d \exp\left\{ b \log\left(\frac{\theta}{1-\theta} \right) \right\} \\
&= (1-\theta)^{nd-b} \theta^b,
\end{aligned}
$$

which is in the Beta family.

Conjugate priors

Some standard prior-likelihood conjugate pairs:

Likelihood	Prior	Posterior
$X \sim \mathsf{Bin}(n, \theta)$	$\mathsf{Beta}(p, q)$	$\mathsf{Beta}(p + x, q + n - x)$
$X_1, \ldots, X_n \sim \mathsf{Poi}(\theta)$	$\mathsf{Gam}(p, q)$	$\mathsf{Gam}(p + \sum_{i=1}^n x_i, q + n)$
$X_1, \ldots, X_n \sim \mathsf{N}(\theta, \tau^{-1})$ (τ known)	$\mathsf{N}(b, c^{-1})$	$\mathsf{N}(\frac{cb+n\tau\bar{x}}{c+n\tau}, \frac{1}{c+n\tau})$
$X_1, \ldots, X_n \sim \mathsf{Gam}(k, \theta)$ (k known)	$\mathsf{Gam}(p, q)$,	$\mathsf{Gam}(p + nk, q + \sum_{i=1}^n x_i)$
$X_1, \ldots, X_n \sim \mathsf{Geo}(\theta)$	$\mathsf{Beta}(p, q)$	$\mathsf{Beta}(p + n, q + \sum_{i=1}^n x_i - n)$
$X \sim \mathsf{NegBin}(r, \theta)$	$\mathsf{Beta}(p, q)$	$\mathsf{Beta}(p + r, q + x)$

You can find more on the Conjugate Prior wikipedia page:
https://en.wikipedia.org/wiki/Conjugate_prior

Outline

1. Recap
2. Different types of priors
    - Conjugate priors, improper priors, Jeffreys' priors
3. Monte Carlo Methods
    - Inversion sampling.

Improper priors

Improper prior definition:

A prior distribution, $\pi(\theta)$, such that $\int \pi(\theta)\mathrm{d}\theta = \infty$.

Example: Normal model

Suppose that $X_1, \ldots, X_n \sim \mathsf{N}(\theta, \sigma^2)$ with σ^2 fixed/known. Then

$$L(x \mid \theta) \propto \exp\left\{-\frac{\sum(x_i - \theta)^2}{2\sigma^2}\right\}.$$

Now with a $\theta \sim \mathsf{N}(b, d^2)$ prior, the posterior becomes:

$$\begin{aligned}
\pi(\theta \mid x) &\propto \exp\left\{-\frac{(\theta - b)^2}{2d^2}\right\} \exp\left\{-\frac{\sum_{i=1}^n (x_i - \theta)^2}{2\sigma^2}\right\} \\
&= \exp\left\{-\frac{(\theta - b)^2}{2d^2} - \frac{\sum_{i=1}^n (x_i - \theta)^2}{2\sigma^2}\right\} \\
&= \exp\left\{-\frac{\theta^2 - 2b\theta + b^2}{2d^2} - \frac{\sum_{i=1}^n x_i^2 - 2n\bar{x}\theta + n\theta^2}{2\sigma^2}\right\}
\end{aligned}$$

Improper priors

$$
\begin{aligned}
\pi(\theta \mid x) &\propto \exp\left\{-\frac{\theta^2-2b\theta+b^2}{2d^2} - \frac{\sum_{i=1}^n x_i^2 - 2n\bar{x}\theta + n\theta^2}{2\sigma^2}\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left[\theta^2\left(\frac{1}{d^2}+\frac{n}{\sigma^2}\right) - 2\theta\left(\frac{b}{d^2}+\frac{n\bar{x}}{\sigma^2}\right)\right]\right\} \\
&\propto \exp\left\{-\frac{1}{2}\left(\frac{1}{d^2}+\frac{n}{\sigma^2}\right)\left(\theta - \frac{\frac{b}{d^2}+\frac{n\bar{x}}{\sigma^2}}{\frac{1}{d^2}+\frac{n}{\sigma^2}}\right)^2\right\}
\end{aligned}
$$

and so

$$
\theta \mid x \sim \mathsf{N}\left(\frac{\frac{b}{d^2}+\frac{n\bar{x}}{\sigma^2}}{\frac{1}{d^2}+\frac{n}{\sigma^2}}, \frac{1}{\frac{1}{d^2}+\frac{n}{\sigma^2}}\right).
$$

Can express more neatly in terms of precisions $\tau = 1/\sigma^2$ and $c = 1/d^2$:

$$
\theta \mid x \sim \mathsf{N}\left(\frac{cb+n\tau\bar{x}}{c+n\tau}, \frac{1}{c+n\tau}\right).
$$

Improper priors

$$\theta \mid x \sim \mathsf{N}\left(\frac{cb + n\tau\bar{x}}{c + n\tau}, \frac{1}{c + n\tau}\right).$$

1. Note that (here $\bar{x}_n := \bar{x}$ reminds us that this is an average of n samples):
$$\mathbb{E}[\theta \mid x] = \gamma_n b + (1 - \gamma_n)\bar{x}_n$$
where
$$\gamma_n = \frac{c}{c + n\tau}.$$
   - $\Rightarrow$ posterior mean is weighted average of prior mean and $\bar{x}$.
   - Weight γ_n is determined by relative strength of information in prior and data components.
   - If $n\tau$ is large relative to c, then $\gamma_n \approx 0$ and posterior mean $\approx \bar{x}$.
2. Posterior precision = prior precision + $n\times$ data precision

Improper priors

$$\theta \,|\, x \sim \mathsf{N}\left(\frac{cb + n\tau\bar{x}_n}{c + n\tau}, \frac{1}{c + n\tau}\right).$$

3. The posterior depends on the data only through $\bar{x}$, and not through the x_i. We say that $\bar{x}$ is sufficient for θ.
4. As $n \to \infty$, then

   $$\theta \,|\, x \sim \mathsf{N}\left(\bar{x}_n, \frac{\sigma^2}{n}\right).$$

   so the the prior has no effect as the amount of data gets large.
5. As prior standard deviation $(d \to \infty)$ (i.e. $c \to 0$), then

   $$\theta \,|\, x \sim \mathsf{N}\left(\bar{x}_n, \frac{\sigma^2}{n}\right).$$

Improper priors

5. As prior standard deviation ($d \to \infty$) (i.e. $c \to 0$), then

$$\theta \,|\, x \sim \mathsf{N}\left(\bar{x}_n, \frac{\sigma^2}{n}\right).$$

- Strength of prior belief is determined by c.
- Small c means weak prior information (i.e. very large variance).
- $c \to 0$ gives the perfectly valid posterior
$$\theta \,|\, x \sim \mathsf{N}\left(\bar{x}_n, \frac{\sigma^2}{n}\right).$$
- But $c \to 0$ is a $\mathsf{N}(0, \infty)$ prior distribution, which has density $\pi(\theta) \propto 1$. Completely uninformative (i.e. uniform on $\mathbb{R}$).
- This is an improper density as $\int_{\mathbb{R}} \pi(\theta) \mathrm{d}\theta = \infty$.
- OK to use, as long as $\pi(\theta \,|\, x)$ is proper.
(Though see Bayes Factors later on.)
- If $\int \pi(\theta \,|\, x) \mathrm{d}\theta = \infty$ then cannot use this type of prior.

Outline

1. Recap
2. Different types of priors
    - Conjugate priors, improper priors, Jeffrey's priors
3. Monte Carlo Methods
    - Inversion sampling.

Problems with improper priors

Previous: representing absolute uninformativeness led to improper priors. However, this can produce problems.

Suppose we specify $\theta \sim \pi(\theta) \propto 1$.

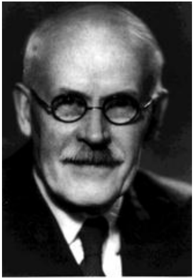
Now consider the parameter $\phi = \theta^2$. Using a change of variable on the prior this gives us

$$\begin{aligned} \pi(\phi) &= \pi(\theta) \times \left| \frac{d\theta}{d\phi} \right| \\ &\propto 1/\sqrt{\phi}. \end{aligned}$$

But: if we are ignorant about θ, then we are equally ignorant about ϕ. So we might equally have stated $\pi(\phi) \propto 1$.

Thus prior ignorance (as represented as uniformity of belief) does not translate across scales.

Jeffreys' Prior



Harold Jeffreys (1891–1989)

Jeffreys' perspective: specification of prior ignorance should be consistent across 1-1 transformations

Jeffreys' prior defined as:

$$\pi_J(\boldsymbol{\theta}) \propto |\mathbf{I}(\boldsymbol{\theta})|^{1/2}$$

where $\mathbf{I}(\boldsymbol{\theta}) = -\mathbb{E}\left[\frac{\mathrm{d}^2 \log L(\boldsymbol{x} \mid \boldsymbol{\theta})}{\mathrm{d}\boldsymbol{\theta}^2}\right]$ is the Fisher information matrix.

Verification for $\boldsymbol{\theta} \in \mathbb{R}$:

Suppose $\phi = g(\theta)$ is a 1-1 transformation of θ. Then

$$\begin{aligned} \pi_J(\phi) &= \pi_J(\theta) \times \left|\frac{\mathrm{d}\theta}{\mathrm{d}\phi}\right| \\ &\propto |I(\theta)|^{1/2} \times \left|\frac{\mathrm{d}\theta}{\mathrm{d}\phi}\right| \\ &= \left| I(\theta) \left(\frac{\mathrm{d}\theta}{\mathrm{d}\phi}\right)^2 \right|^{1/2} \end{aligned}$$

Jeffreys' Prior

$$
\begin{aligned}
&= \left| I(\theta) \left(\frac{\mathrm{d}\theta}{\mathrm{d}\phi} \right)^2 \right|^{1/2} \\
&= \left| -\mathbb{E}\left[\left(\frac{\mathrm{d}\log L}{\mathrm{d}\theta} \right)^2 \right] \left(\frac{\mathrm{d}\theta}{\mathrm{d}\phi} \right)^2 \right|^{1/2} \\
&= \left| -\mathbb{E}\left[\left(\frac{\mathrm{d}\log L}{\mathrm{d}\theta} \right)^2 \left(\frac{\mathrm{d}\theta}{\mathrm{d}\phi} \right)^2 \right] \right|^{1/2} \\
&= \left| -\mathbb{E}\left[\left(\frac{\mathrm{d}\log L}{\mathrm{d}\theta} \frac{\mathrm{d}\theta}{\mathrm{d}\phi} \right)^2 \right] \right|^{1/2} \\
&= \left| -\mathbb{E}\left[\left(\frac{\mathrm{d}\log L}{\mathrm{d}\phi} \right)^2 \right] \right|^{1/2} = |I(\phi)|^{1/2}
\end{aligned}
$$

as required.

Jeffreys' Prior

Example: Binomial model

Suppose $X \mid \theta \sim \mathsf{Bin}(n, \theta)$.

Then

$$\log(L(x \mid \theta)) = x \log(\theta) + (n - x) \log(1 - \theta),$$

so

$$\frac{\mathrm{d}^2 \log L(x \mid \theta)}{\mathrm{d}\theta^2} = -\frac{x}{\theta^2} - \frac{(n - x)}{(1 - \theta)^2}$$

and since $\mathbb{E}X = n\theta$,

$$\begin{aligned} I(\theta) \quad &= \tfrac{n\theta}{\theta^2} + \tfrac{n - n\theta}{(1-\theta)^2} \\ &= \tfrac{n}{\theta} + \tfrac{n}{(1-\theta)} \\ &= n\theta^{-1}(1 - \theta)^{-1}, \end{aligned}$$

leading to

$$\pi_J(\theta) \propto \theta^{-1/2}(1 - \theta)^{-1/2}$$

which in this case is the proper distribution $\mathrm{Beta}(\frac{1}{2}, \frac{1}{2})$.

Jeffreys' Prior

Example: Normal mean model (N_1)

Suppose $X_1, \ldots, X_n$ are *iid* $\mathsf{N}(\theta, \sigma^2)$, ($\sigma^2$ known).

Then

$$L(x \mid \theta) \propto \exp\left(-\frac{\sum_{i=1}^n (x_i - \theta)^2}{2\sigma^2}\right) \quad \text{so} \quad \log(L(x \mid \theta)) = -\frac{\sum_{i=1}^n (x_i - \theta)^2}{2\sigma^2}$$

and

$$\frac{\mathrm{d}^2 \log L(x \mid \theta)}{\mathrm{d}\theta^2} = -\frac{n}{\sigma^2}$$

$$\begin{aligned} I(\theta) &= -\mathbb{E}\left[\frac{\mathrm{d}^2 \log L(x \mid \theta)}{\mathrm{d}\theta^2}\right] \\ &= \mathbb{E}\frac{n}{\sigma^2} = \frac{n}{\sigma^2} \end{aligned}$$

Hence

$$\pi_J(\theta) \propto 1$$

the improper uniform distribution is Jeffrey's prior.

Jeffreys' Prior

Example: Normal variance model (N_2)

Suppose $X_1, \ldots, X_n$ are *iid* $\mathsf{N}(m, \theta)$, (m known), then

$$L(x \,|\, \theta) \propto \theta^{-n/2} \exp\{-s/(2\theta)\} \quad \text{and} \quad \log L(x \,|\, \theta) \propto -\frac{n}{2} \log \theta - \frac{s}{2\theta}$$

where $s = \sum_i (x_i - m)^2$.

Then

$$\frac{\partial^2 \log L(x \,|\, \theta)}{\partial \theta^2} = \left(\frac{n}{2\theta^2}\right) - \left(\frac{s}{\theta^3}\right)$$

and since $\mathbb{E}[s \,|\, \theta] = n\theta$, Jeffreys' prior is

$$\pi_J(\theta) \propto \theta^{-1}$$

.

Jeffreys' Prior

Example: Normal variance model (N_3)

Suppose $X_1, \ldots, X_n$ are *iid* $\mathsf{N}(\mu, \sigma^2)$ with $\boldsymbol{\theta} = (\mu, \sigma^2)$ then

$$L(x \mid \theta) \propto \sigma^{-n} \exp\left[-\frac{n}{2\sigma^2}\{s + (\bar{x} - \mu)^2\}\right],$$

where $s = n^{-1} \sum_i (x_i - \bar{x})^2$. Then

$$\begin{aligned} \mathbf{I}(\theta) &= \mathbb{E}\begin{pmatrix} n/\sigma^2 & n(\bar{x} - \mu)/\sigma^4 \\ n(\bar{x} - \mu)/\sigma^4 & -\{n/(2\sigma^4)\} + n\{s + (\bar{x} - \mu)^2\}/\sigma^6 \end{pmatrix} \\ &= \begin{pmatrix} n/\sigma^2 & 0 \\ 0 & n/(2\sigma^4) \end{pmatrix}. \end{aligned}$$

And so Jeffreys prior is $\pi_J(\mu, \sigma^2) \propto \sigma^{-3}$.

Objections to Jeffreys' Prior

Objection 1:

Jeffreys' prior depends on the form of the data (through $L(x \mid \theta)$). A prior distribution should not be influenced by what data are to be collected.

Objection 2:

In certain cases, Jeffreys' prior is inconsistent. E.g.

- N_1: $x \sim \mathsf{N}(\theta, \sigma^2)$ (σ known), $\pi_J(\theta) \propto 1$
- N_2: $x \sim \mathsf{N}(m, \theta)$ (m known), $\pi_J(\theta) \propto \theta^{-1}$
- N_3: $x \sim \mathsf{N}(\mu, \sigma^2)$, $\pi_J(\theta) \propto \sigma^{-3}$ both unknown

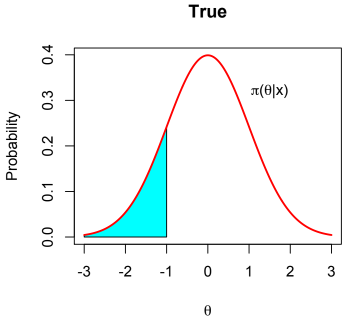
Here $N_3 \neq N_2 \times N_1$.

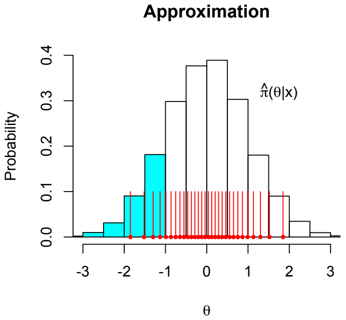
Jeffreys argues that in this case ignorance about μ and σ^2 should be represented by independent ignorance priors for the two parameters separately.

Outline

1. Recap
2. Different types of priors
   - Conjugate priors, improper priors, Jeffrey's priors
3. Monte Carlo Methods
   - Inversion sampling.

Recap: Posterior simulation





$$\begin{aligned}\mathbb{P}[\theta < -1] &= \\ \int_\theta \mathbb{I}[\theta < -1]\pi(\theta \,|\, x)\mathrm{d}\theta &= \\ \mathbb{E}_\pi \mathbb{I}[\theta < -1]\end{aligned}$$

$$\mathbb{P}[\theta < -1] \approx \frac{1}{N}\sum_{i=1}^{N}\mathbb{I}[\theta^{(i)} < -1]$$

where $\theta^{(1)}, \ldots, \theta^{(N)} \sim \pi(\theta \,|\, x)$.

- We are interesting in ways to draw samples from distributions.
- ∃ many such Monte Carlo algorithms – we will look at several.
- This week: Inversion sampling (the simplest)

Inversion Sampling

Problem: How can we simulate from a univariate distribution $F(x)$?

Application of the probability integral transformation:

$$\text{If } X \sim F(x), \text{ then } F(X) \sim U(0, 1).$$

Inversion Sampling

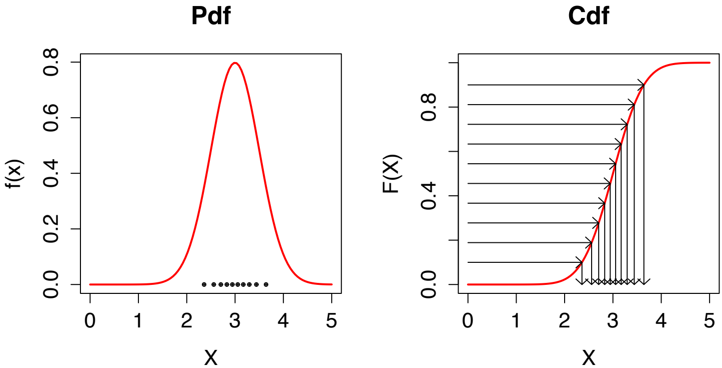
Inverting the probability integral transformation:

$$\text{If } \mathsf{U} \sim U(0, 1) \text{ then } F^{-1}(U) = X.$$

Hence, to draw n samples from $F(x)$:

- Draw $U_1, \ldots, U_n \sim \mathsf{U}(0, 1)$
- Set $X_i = F^{-1}(U_i)$.
- Return $X_1, \ldots, X_n$ as *iid* samples from $F(x)$.

Inversion Sampling: Why does it work?



Let $X = F^{-1}(U)$ where U is uniformly distributed on $[0, 1]$.
The distribution function of X is:

$$\begin{aligned}
\mathbb{P}(X \leq x) &= \mathbb{P}(F^{-1}(U) \leq x) \\
&= \mathbb{P}(U \leq F(x)) \\
&= \int_0^{F(x)} 1 \mathrm{d}u \\
&= F(x).
\end{aligned}$$

Example: Inversion Sampling

Problem: Devise a method for simulating a r. v. with the density

$$f(x) = 3x^2 \text{ for } x \in (0,1).$$

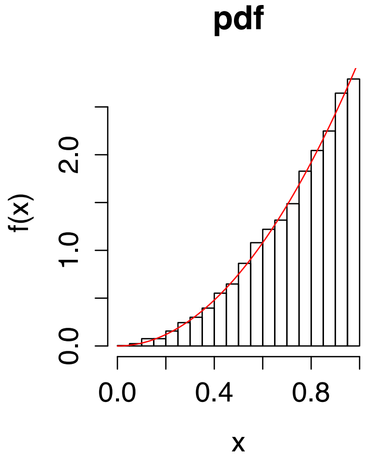
Solution:

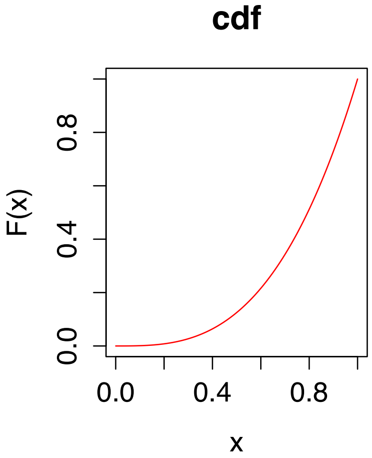
First compute the distribution function: $F(x) = x^3$.

Write $F(x) = u$ and then solve for x: $x = u^{1/3}$.

i.e. Generate $U \sim \mathsf{U}(0,1)$ and set $X = U^{1/3} \sim f(x)$.

Example: Inversion Sampling






```
x=runif(5000)^(1/3)
hist(x,probability=T)
```

Example: Inversion Sampling

Problem: Devise a method for simulating a r. v. with the density

$$f(x) = \begin{cases} \frac{(x-2)}{2} & \text{if } 2 \le x \le 3 \\ \frac{(2-x/3)}{2} & \text{if } 3 < x \le 6 \\ 0 & \text{otherwise.} \end{cases}$$

Solution:
First compute the distribution function:
If $x < 2$, $F(x) = 0$.
If $2 \le x \le 3$, then

$$\begin{aligned} F(x) &= \int_2^x \left(\frac{z-2}{2}\right) dz \\ &= \left[\frac{(z-2)^2}{4}\right]_2^x = \frac{(x-2)^2}{4}. \end{aligned}$$

Example: Inversion Sampling

If $3 < x \le 6$, then

$$\begin{aligned} F(x) &= F(3) + \int_3^x \frac{2 - z/3}{2} dz \\ &= \frac{1}{4} + \left[-\frac{3(2 - z/3)^2}{4} \right]_3^x \\ &= \frac{1}{4} + \left[-\frac{3(2 - x/3)^2}{4} + \frac{3}{4} \right] \\ &= 1 - \frac{3}{4}(2 - x/3)^2. \end{aligned}$$

If $x > 6$, then $F(x) = 1$.

Now compute $x = F^{-1}(u)$.
When $x \in (2, 3]$ (equivalently $u \in (F(2), F(3)] = (0, 1/4]$) we have

$$\begin{aligned} (x-2)^2/4 &= u \\ x - 2 &= \sqrt{4u} \\ x &= 2 + 2\sqrt{u} = F^{-1}(u). \end{aligned}$$

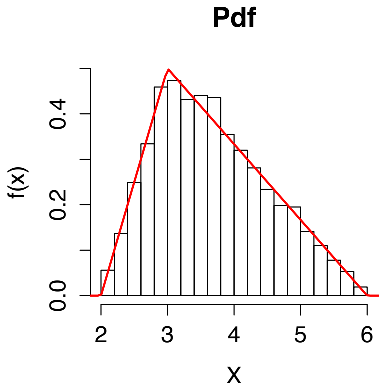
Example: Inversion Sampling

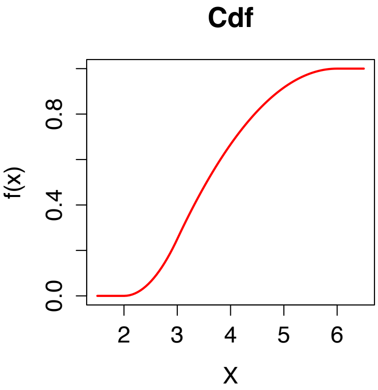
When $x \in (3,6]$ (equivalently $u \in (F(3), F(6)] = (1/4, 1]$) we have

$$\begin{aligned}
1 - 3/4(2 - x/3)^2 &= u \\
1 - u &= 3/4(2 - x/3)^2 \\
4/3(1-u) &= (2 - x/3)^2 \\
\sqrt{4/3(1-u)} &= 2 - x/3 \\
x/3 &= 2 - \sqrt{4/3(1-u)} \\
x &= 3(2 - \sqrt{4/3(1-u)}) = F^{-1}(u).
\end{aligned}$$

$$F^{-1}(u) = \begin{cases} 2 + \sqrt{4u} & u \in (0, 1/4] \\ 3(2 - \sqrt{4/3(1-u)}) & u \in (1/4, 1) \end{cases}$$

Example: Inversion Sampling






```
Finv=function(u) {
  out=rep(0,length(u))
  ind=(u<=0.25)
  out[ind] = 2+2*sqrt(u[ind])
  ind=(u>0.25)
  out[ind] = 6*(1-sqrt((1-u[ind])/3))
  return(out)
}
```

```
x=Finv(runif(5000))
hist(x,probability=T)
```