read E&R Chapter 7 & Section 3.5

VI Bayesian Inference

1 Basic Concepts

- we have the model for the data x

$$\mathcal{M} = \{f_\theta : \theta \in \Theta\}$$

| one value of θ is true.

- think of f_θ as the conditional (prob.) dist. of x given θ

- to this we add a prob. measure for θ as given by density π called the <u>prior</u>

- is θ random?

- recall: probability measures beliefs among what is true. so

$$\Pi(A) = \sum_{\theta \in A} \pi(\theta)$$

= prior prob. that true value of θ is in $A \subseteq \Theta$

- so prior and model lead to a joint probability distribution for (θ, x) given by using multiplication rule

$$\pi(\theta) f_\theta(x)$$

- we observe x

- <u>principle of conditional probability</u>

  Suppose we have a prob. measure P on subsets of $\underline{\Omega}$ and we make the assignment $P(A)$ to $A \subseteq \underline{\Omega}$, i.e. our degree of belief that an unknown $\omega \in \underline{\Omega}$ is in A is $P(A)$. We observe the value i_0 of a function $i : \underline{\Omega} \to \mathcal{I}$ so we know that $\omega \in i^{-1}\{i_0\} = C$. Then we replace $P(A)$ by

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

  as our measure of belief that $\omega \in A$.

- a basic principle (axiom) of inference.

eg - rolling a symmetric die

  - $\underline{\Omega} = \{1,2,3,4,5,6\}$

  - $A = \{2,3,6\}$, $P(A) = \frac{1}{2}$

  - told "outcome is even"

  - $C = \{2,4,6\}$

$$P(A|C) = \frac{P(A \cap C)}{P(C)} = \frac{P(\{2,6\})}{P(\{2,4,6\})}$$

$$= 2/3$$ belief goes up

- add "outcome is divisible by 3"
  so $C = \{3,6\}$

$$P(A|C) = \frac{P(\{3,6\})}{P(\{3,6\})} = 1$$

- add "outcome is odd" so $C = \{1,3,5\}$

$$P(A|C) = \frac{P(\{3\})}{P(\{1,3,5\})} = \frac{1}{3}$$ belief goes down

- <u>note</u> be careful - need to specify i (information generator)

- in the statistical problem $i(\theta, x) = x$

- principle of conditional probability says that we replace the prior π by the <u>posterior</u> distribution.

$$\pi(\theta|x) = \frac{\pi(\theta) f_\theta(x)}{m(x)}$$

where $m(x) = \begin{cases} \sum_{\theta \in \Theta} \pi(\theta) f_\theta(x) & \text{discrete} \\ \int_\Theta \pi(\theta) f_\theta(x)\, dx & \text{cnt} \end{cases}$

is the marginal distribution of x
called the __prior predictive__ of x

- so $\pi(A)$ is replaced by

$$\pi(A|x) = \sum_{\theta \in A} \pi(\theta|x)$$

- ~~the~~ principle of conditional prob says we replace the prior by ~~the~~ posterior for prob statements about Θ.

- what about prob. statements about $\tau = \Upsilon(\theta)$?

- marginal prior density $\pi_\Upsilon(\tau) = \sum_{\theta \in \Upsilon^{-1}\{\tau\}} \pi(\theta)$

marginal posterior density $\pi_\Upsilon(\tau|x) = \sum_{\theta \in \Upsilon^{-1}\{\tau\}} \pi(\theta|x)$

- prediction: suppose we have an unobserved y with conditional model $\{g_{\beta(\theta)}(\cdot|x) : \theta \in \Theta\}$ where the same value of θ is true

eg x = marks on assignments

y = mark on final.

- the prior predictive for y

$$m_Y(y) = \sum_{\Theta}\sum_{x} \pi(\theta) f_\theta(x) g_{\beta(\theta)}(y|x)$$

the posterior predictive for y

$$m_Y(y|x) = \frac{\text{joint for } (x,y)}{\text{marginal for } x}$$

(3 variables θ, x, y)

$$= \frac{\sum_{\Theta} \pi(\theta) f_\theta(x) g_{\beta(\theta)}(y|x)}{\sum_{\Theta}\sum_{y} \pi(\theta) f_\theta(x) g_{\beta(\theta)}(y|x)}$$

$$= \frac{m(x) \sum_{\Theta} \pi(\theta|x) g_{\beta(\theta)}(y|x)}{m(x) \sum_{\Theta} \pi(\theta) f_\theta(x) \sum_{y} g_{\beta(\theta)}(y|x)}$$

($\sum_{y} g_{\beta(\theta)}(y|x) = 1$)

$$= \sum_{\Theta} \pi(\theta|x) g_{\beta(\theta)}(y|x)$$

(7) Prior Elicitation

- how do we choose π in a problem?
- note - fair to ask - how do we choose the model M in a problem?
  - "we choose" it - it is subjective
- we base the choice of π on what we know about the measurements we are taking

eg location-scale normal

- basic measurement $x_i \sim N(\mu, \sigma^2)$ where $\mu \in \mathbb{R}^1$, $\sigma^2 > 0$ are unknown
- common choice (specify π hierarchically)

$$\mu \mid \sigma^2 \sim N(\mu_0, \tau_0^2 \sigma^2)$$

$$\frac{1}{\sigma^2} \sim \text{gamma}_{\text{rate}}(\alpha_0, \beta_0)$$

where $\mu_0, \tau_0^2, \alpha_0, \beta_0$ are chosen "hyperparameters"

- joint for sample $((\mu, 1/\sigma^2), x)$ where $x = (x_1, \ldots, x_n)$

$$(2\pi\sigma^2)^{-\frac{n}{2}} \exp\left\{-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2\right\} \exp\left\{-\frac{(n-1)s_x^2}{2\sigma^2}\right\} \times$$

$$(2\pi\tau_0^2\sigma^2)^{-\frac{1}{2}} \exp\left\{-\frac{1}{2\tau_0^2\sigma^2}(\mu-\mu_0)^2\right\} \times$$

$$\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} \left(\frac{1}{\sigma^2}\right)^{\alpha_0-1} \exp\{-\beta_0/\sigma^2\}$$

<u>posterior</u>

$$\propto \exp\left\{-\frac{1}{2\sigma^2}\left[n(\bar{x}-\mu)^2+\frac{1}{\tau_0^2}(\mu-\mu_0)^2\right]\right\} \times$$

$$\left(\frac{1}{\sigma^2}\right)^{\frac{n+1+2\alpha_0}{2}-1} \exp\left\{-\left[\frac{(n-1)s_x^2}{2}+\beta_0\right]/\sigma^2\right\}$$

$$n(\bar{x}-\mu)^2+\frac{1}{\tau_0^2}(\mu-\mu_0)^2$$

$$= n\bar{x}^2 - 2n\bar{x}\mu + n\mu^2 + \frac{1}{\tau_0^2}\mu^2 - \frac{2\mu_0}{\tau_0^2}\mu + \frac{\mu_0^2}{\tau_0^2}$$

$$= \left(n+\frac{1}{\tau_0^2}\right)\mu^2 - 2\left(n\bar{x}+\frac{\mu_0}{\tau_0^2}\right)\mu + n\bar{x}^2 + \frac{\mu_0^2}{\tau_0^2}$$

$$= \left(n+\frac{1}{\tau_0^2}\right)\left(\mu - \left(n+\frac{1}{\tau_0^2}\right)^{-1}\left(n\bar{x}+\frac{\mu_0}{\tau_0^2}\right)\right)^2$$

$$- \left(n+\frac{1}{\tau_0^2}\right)^{-1}\left(n\bar{x}+\frac{\mu_0}{\tau_0^2}\right)^2 + n\bar{x}^2 + \frac{\mu_0^2}{\tau_0^2}$$

$$= (\quad) - \left(n+\frac{1}{\tau_0^2}\right)^{-1}\left(n^2\bar{x}^2 + \frac{2n\bar{x}\mu_0}{\tau_0^2} + \frac{\mu_0^2}{\tau_0^4}\right)$$

$$+ n\bar{x}^2 + \frac{\mu_0^2}{\tau_0^2}$$

$$= (\;) + \left(n - \frac{n^2\tau_0^2}{n\tau_0^2+1}\right)\bar{x}^2 - \frac{2n\tau_0^2}{n\tau_0^2+1}\bar{x}\mu_0 + \left(\frac{1}{\tau_0^2} - \frac{1}{n\tau_0^2+1}\frac{1}{\tau_0^2}\right)$$

$$= (\ \ ) + \frac{n}{n\tau_0^2+1}\bar{x}^2 - \frac{2n}{n\tau_0^2+1}\bar{x}\mu_0 + \frac{n\tau_0^2}{n\tau_0^2+1}\frac{1}{\tau_0^2}\mu_0^2$$

$$= (\ \ ) + \frac{n}{n\tau_0^2+1}(\bar{x}-\mu_0)^2$$

$$\propto \left(\frac{1}{\sigma^2}\right)^{\frac{1}{2}} \exp\left\{-\frac{1}{2\sigma^2}\left(n+\frac{1}{\tau_0^2}\right)\left(\mu-\left(n+\frac{1}{\tau_0^2}\right)^{-1}\left(n\bar{x}+\frac{\mu_0}{\tau_0^2}\right)\right)^2\right\}$$

$$\times \left(\frac{1}{\sigma^2}\right)^{\frac{n+2\alpha_0}{2}-1} \exp\left\{-\left[\frac{(n-1)s^2}{2}+\beta_0+\frac{1}{2}\frac{n}{n\tau_0^2+1}(\bar{x}-\mu_0)^2\right]\Big/\frac{1}{\sigma^2}\right\}$$

$$\therefore\ \mu \mid x, \sigma^2 \sim N\left(\left(n+\frac{1}{\tau_0^2}\right)^{-1}\left(n\bar{x}+\frac{\mu_0}{\tau_0^2}\right), \left(n+\frac{1}{\tau_0^2}\right)^{-1}\sigma^2\right)$$

$$\frac{1}{\sigma^2}\Big| x \sim \text{gamma}_{\text{rate}}\left(\frac{n+2\alpha_0}{2}, \underbrace{\beta_0+\frac{(n-1)s_x^2}{2}+\frac{1}{2}\frac{n}{n\tau_0^2+1}(\bar{x}-\mu_0)^2}_{\beta_x}\right)$$

<u>note</u> – the posterior is of the same form as the prior

– the prior is said to be <u>conjugate</u>.

<u>note</u> can generate from the prior and posterior by first generating $1/\sigma^2$ and then generating $\mu \mid \sigma^2$

– still how do we elicit $\mu_0, \tau_0^2, \alpha_0, \beta_0$?

<u>elicitation</u>

- suppose we are measuring heights of students at U of T (in some units) so μ = mean height , σ^2 = variance of hts

- specify (m_1, m_2) so that you believe that $\mu \in (m_1, m_2)$ with virtual certainty (say prob $\geq .99$) and put $\mu_0 = (m_1+m_2)/2$

- so

$$0.99 \leq \Phi\left(\frac{m_2-\mu_0}{\tau_0\sigma}\right) - \Phi\left(\frac{m_1-\mu_0}{\tau_0\sigma}\right)$$
$$= \Phi\left((m_2-m_1)/2\tau_0\sigma\right) - \Phi\left(-(m_2-m_1)/2\tau_0\sigma\right)$$
$$= 2\Phi\left(\frac{m_2-m_1}{2\tau_0\sigma}\right) - 1$$

iff $\frac{1+0.99}{2} = 0.995 \leq \Phi\left(\frac{m_2-m_1}{2\tau_0\sigma}\right)$

iff $z_{0.995} = \Phi^{-1}(0.995) \leq \frac{m_2-m_1}{2\tau_0\sigma}$ | $z_{0.995} = 2.576$

iff $\sigma \leq \frac{m_2-m_1}{2\tau_0 z_{0.995}}$ (*) | haven't prescribed τ_0 yet

- we know a measurement lies in $(\mu - \sigma z_{0.995}, \mu + \sigma z_{0.995})$ with virtual certainty

- let s_1, s_2 be lower and upper bounds on the half-length of this interval

$s_1 \leq \sigma z_{0.995} \leq s_2$ or $\frac{z^2_{0.995}}{s_2^2} \leq \frac{1}{\sigma^2} \leq \frac{z^2_{0.995}}{s_1^2}$

- combining this with ⊛ gives.

$$\frac{m_2 - m_1}{2\tau_0 z_{0.995}} = \frac{s_2}{z_{0.995}}$$

or $\tau_0 = \frac{m_2 - m_1}{2 s_2}$ so (μ_0, τ_0) is determined

- need to determine (α_0, β_0)
- let $G(\alpha, \beta, \cdot)$ denote the gamma(α,β) cdf with quantile fn $G^{-1}(\alpha, \beta, \cdot)$, note $G(\alpha, \beta, x) = G(\alpha, 1, \beta x)$
- so we want α, β s.t.

$$G\left(\alpha, \beta, \frac{z^2_{0.995}}{s_1^2}\right) = 0.995$$

$$G\left(\alpha, \beta, \frac{z^2_{0.995}}{s_2^2}\right) = 0.005$$

- do this iteratively pick α_0 and find w_0 s.t. $G^{-1}(\alpha_0, 1, 0.995) = w_0$ so $\beta_0 = w_0 s_1^2 / z^2_{0.995}$
- if $G\left(\alpha_0, 1, \beta_0 \frac{z^2_{0.995}}{s_2^2}\right) \lessgtr 0.005$ then ~~decrease~~ increase α_0 and repeat

- in general: statistical analyses require an elicitation procedure like this

- **note** - if $\pi(\theta) = 0$ then $\pi(\theta|x) = \frac{\pi(\theta) f_\theta(x)}{m(x)} = 0$

  so never put 0 mass at θ unless you categorically know that θ is false.

(3) Checking the Ingredients

- we have chosen a model $\{f_\theta : \theta \in \Theta\}$ and a prior π
- do these choices make sense?
- how do we assess this?
- principle of empirical criticism

  Any ingredient chosen to be part of a statistical analysis must be checked against the objective data to see if there is any indication it is a bad choice.

- **note** - we can never say a choice is correct only that it is not clearly wrong and so potentially misleading

- **note** - an important corollary of the principle: if inferences depend on ingredients that cannot be checked against the data then the analysis must be so qualified

- **note** - we want to check the model first and then the prior

- how?
- let T be a mss for the model
- we have the joint dist of (θ, x)

$$\pi(\theta) f_\theta(x) = \pi(\theta) f_{\theta,T}(T(x))\, f(x \mid T(x))$$

$$= \pi(\theta \mid T(x))\, m_T(T(x))\, f(x \mid T(x))$$

- <u>note</u> $\pi(\theta \mid x) = \dfrac{\pi(\theta) f_\theta(x)}{m(x)} = \dfrac{\pi(\theta) f_{\theta,T}(T(x)) f(x \mid T(x))}{\sum_\theta f_{\theta,T}(T(x)) f(x \mid T(x)) \pi(\theta)}$

$$= \frac{\pi(\theta) f_{\theta,T}(T(x))}{m_T(T(x))}$$

$m(x) = m_T(T(x))\, f(x \mid T(x))$

so posterior only depends on data through T.

- joint factors into $\pi(\theta \mid T(x))$ probs for θ

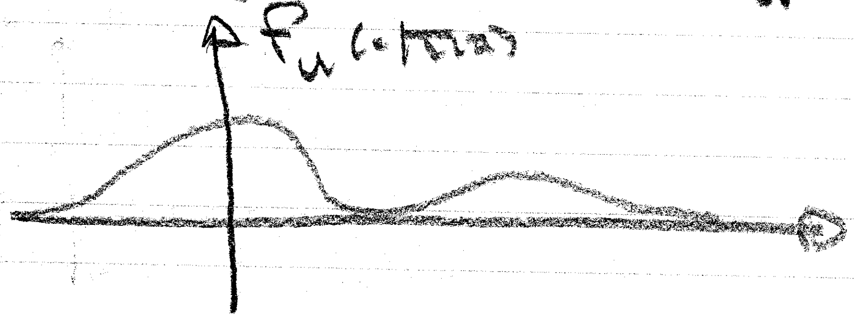
$m_T(T(x))$ checking the prior

$f(x \mid T(x))$ checking the model

checking the model

- let $U(x)$ be a discrepancy statistic in the sense that "large" values of $U(x)$ indicate model failure.
- then compute

$$P\left(f_U(u \mid T(x)) \le f_U(U(x) \mid T(x)) \right)$$



- a small tail prob indicates problem with the model.

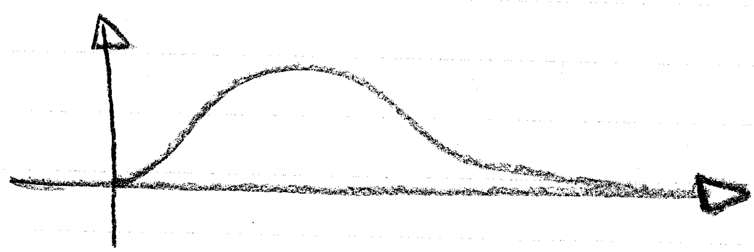
Eg $\underline{x} = (x_1, \ldots, x_n) \overset{iid}{\sim} N(\mu, \sigma_0^2)$, $\underline{x} \sim N_n(\mu \underline{1}, \sigma_0^2 I)$

- $T(x) = \bar{x}$
- $\underline{x} = \bar{x}\underline{1} + (\underline{x} - \bar{x}\underline{1})$
- $\underline{x} - \bar{x}\underline{1} = (I - \underline{1}(\underline{1}'\underline{1})^{-1}\underline{1}')\underline{x}$

$$\sim N_n(\underline{0}, \sigma_0^2 (I - \underline{1}(\underline{1}'\underline{1})^{-1}\underline{1}')) \text{ ind of } \bar{x}$$

- $U(x) = \|\underline{x} - \bar{x}\underline{1}\|^2 \sim \sigma_0^2 \text{ chi-squared}(n-1)$

$\therefore P\left(f_U(u \mid T(x)) \le f_U(u(x) \mid T(x)) \right)$



$= P\left(X^{\frac{n-1}{2}-1} e^{-\frac{x}{2\sigma^2}} \le \left(\|x - \bar{x}1\|^2\right)^{\frac{n-1}{2}-1} e^{-\frac{\|x-\bar{x}1\|^2}{2\sigma^2}} \right)$

and compute using simulation.

Eg $x = (x_1, \ldots, x_n)' \sim N(\mu, \sigma^2)$ $\mu \in R^1, \sigma^2 > 0$

- $x \sim N_n(\mu 1, \sigma^2 I)$
- $x = \bar{x}1 + \|x - \bar{x}1\| \left(\frac{x - \bar{x}1}{\|x - \bar{x}1\|} \right)$
- $\bar{x} \sim N(\mu, \sigma^2/n)$

  $\|x - \bar{x}1\|^2 \sim \sigma^2$ Chisquared$(n-1)$ — independent

  $\frac{x - \bar{x}1}{\|x - \bar{x}1\|} \sim U(S^n \cap \mathcal{L}^\perp(1))$
- $U(x) = \frac{1}{n} \sum_{i=1}^{n} \frac{(x_i - \bar{x})^3}{\|x - \bar{x}1\|^3}$ skewness

$= U\left(\frac{x - \bar{x}1}{\|x - \bar{x}1\|} \right)$

- need to compute

$$P\left(f_U(u \mid T(x)) \le f_U(U(x) \mid T(x)) \mid T(x) \right)$$

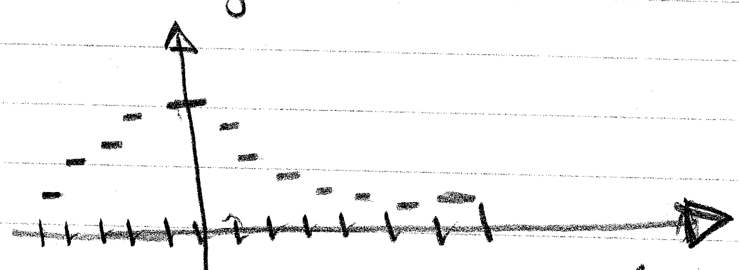
(recall U is ind of T)

$$= P\left(f_U(u) \le f_U(U(x)) \right)$$

algorithm: - generate $\underset{\sim}{z}_1, \ldots, \underset{\sim}{z}_N \overset{iid}{\sim} N_n(\underset{\sim}{0}, I)$

- compute $u_i = U(\underset{\sim}{z}_i)$ $i = 1, \ldots, n$ which gives a sample of n from f_U

- compute estimate $\hat{f}_U$ via a density estimation algorithm, e.g. density histograms

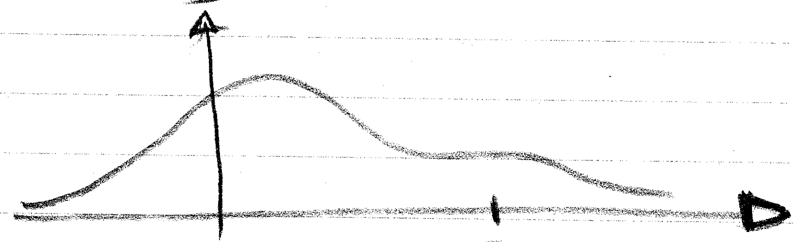


and record proportion $\hat{f}_U(u_i) \le \hat{f}_U(U(x))$

- what to do if the model fails?

checking the prior

- is θ_{true} a plausible value for π?
- measure by $\Pi(\pi(\theta) \le \pi(\theta_{true}))$



- but we don't know θ_{true}!
- Fact - under conditions (T mss)

$$M_T(m_T(t) \le m_T(T(x)))$$

$$\to \Pi(\pi(\theta) \le \pi(\theta_{true}))$$

- ∴ check the prior by computing ⊛
- actually you can check individual components of a prior

— examples

eg ① — $x = (x_1, \ldots, x_n)' \sim N(\mu, 1)$

— $\mu \sim N(\mu_0, \tau_0^2)$

— $T(x) = \bar{x} = \mu + z/\sqrt{n} \overset{\text{prior}}{\sim} N(\mu_0, \tau_0^2 + 1/n)$

— so $m_T(t) = (2\pi)^{-\frac{1}{2}} (\tau_0^2 + \frac{1}{n})^{-\frac{1}{2}} \exp\{-\frac{(\tau_0^2 + 1/n)^{-1}}{2} (t - \mu_0)^2\}$

— $M_T(m_T(t) \le m_T(\bar{x}))$

$= M_T\left((\tau_0^2 + 1/n)^{-1} (t - \mu_0)^2 \ge (\tau_0^2 + 1/n)^{-1} (\bar{x} - \mu_0)^2 \right)$

$= 2\left(1 - \Phi\left((\tau_0^2 + 1/n)^{-1/2} |\bar{x} - \mu_0| \right)\right)$

$\rightarrow \begin{cases} 2\left(1 - \Phi\left(\left| \frac{\mu_{true} - \mu_0}{\tau_0} \right| \right)\right) = \Pi(\pi(\mu) \le \pi(\mu_{true})) & \text{as } n \to \infty \\ 2(1 - \Phi(0)) = 1 & \text{as } \tau_0^2 \to \infty \end{cases}$

— so large τ_0^2 avoids prior-data conflict

eg ② $x = (x_1, \ldots, x_n) \sim \text{Bernoulli}(\theta)$

- $\theta \sim U(0,1)$ (noninformative?)
- $T(x) = n\bar{x}$ is MSS $\sim \text{Binomial}(n, \theta)$
- $m_T(t) = \int_0^1 \binom{n}{t} \theta^t (1-\theta)^{n-t} \, d\theta$

$$= \binom{n}{t} \frac{\Gamma(t+1)\Gamma(n-t+1)}{\Gamma(n+2)} = \frac{n!}{t!(n-t)!} \frac{t!(n-t)!}{(n+1)!} = \frac{1}{n+1}$$

- $M_T(m_T(t) \le m_T(\tilde{t})) = 1 = \Pi(\pi(\theta) \le \pi(\theta_{true}))$

and never any indication of prior-data conflict

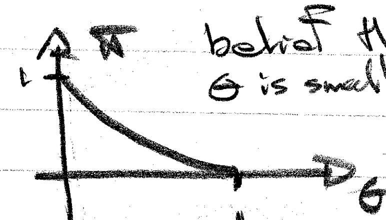
- $\theta \sim \text{beta}(\alpha, \beta)$ $\quad \theta \mid x \sim \text{beta}(n\bar{x}+\alpha, n(1-\bar{x})+\beta)$ so conjugate

$$m_T(t) = \frac{\Gamma(n+1)}{\Gamma(t+1)\Gamma(n-t+1)} \frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)} \frac{\Gamma(n\bar{x}+\alpha)\Gamma(n(1-\bar{x})+\beta)}{\Gamma(n+\alpha+\beta)}$$

eg ③ $y \sim \text{Geometric}(\theta)$ $\quad \theta \in (0,1]$ ← note

- $f_\theta(y) = \theta(1-\theta)^{y-1}$ $\quad y \in \{1, 2, \ldots\}$
- $T(y) = y$ is a MSS (Ex)
- $\theta \sim \text{Beta}(1, c)$ with $c \geq 1$

so $\pi(\theta) = \frac{\Gamma(1+c)}{\Gamma(1)\Gamma(c)} (1-\theta)^{c-1}$



belief that θ is small

- $m_T(t) = c \int_0^1 \theta (1-\theta)^{c+t-2} \, d\theta$ $\quad \Big|\; \begin{matrix} t=1 \\ c=1 \end{matrix}$

$\Gamma(a+1) = a\Gamma(a)$

$$= c\,\frac{\Gamma(2)\,\Gamma(c+t-1)}{\Gamma(c+t+1)} = c\,\frac{(c+t-2)\,\Gamma(c+t-1)}{(c+t)(c+t-1)(c+t-2)\,\Gamma(\;)}$$

- $M_T\,(m_T(t) \le m_T(y))$

$$\ln m_T(t) = \ln c - \ln(c+t) - \ln(c+t-1)$$

$$\frac{\partial}{\partial t} = -\frac{1}{c+t} - \frac{1}{c+t-1} < 0 \quad \text{and}$$

so strictly decreasing in t

$$= \sum_{t=y}^{\infty} \frac{c}{(c+t)(c+t-1)} = -c \sum_{t=y}^{\infty}\left(\frac{1}{c+t} - \frac{1}{c+t-1}\right)$$

$$= -c\left(-\frac{1}{c+y-1}\right) = \frac{c}{c+y-1} = \frac{1}{1+\frac{y-1}{c}}$$

and so we get prior-data conflict when $\left(\frac{y}{c}\right)$ is big as we need lots of tosses to get a H

- **note** - when $c=1$ then $\theta \sim U((0,1])$, P-value is $1/y$

  - why not 1 as with Binomial?

- the bigger c is the more mass we have around 0 but recall $\theta \neq 0$

- in terms of the cont approx this means θ cannot be very small

- a large value of y indicates θ is small and so indicates prior-data conflict

- <u>note</u> - suppose $\theta \in \{\frac{1}{N}, \frac{2}{N}, \ldots, 1\}$ equiprobable

  - $m_T(t) = \frac{1}{N} \sum_{i=1}^{N} \frac{i}{N}(1-\frac{i}{N})^{t-1}$

    $\frac{\partial m_T(t)}{\partial t} = \frac{1}{N} \sum_{i=1}^{N} \left[\ln(1-\frac{i}{N})\right] \frac{i}{N}(1-\frac{i}{N})^{t-1}$

    < 0 so decreasing in t

  - $M_T(m_T(t) \le m_T(y))$

    $= \sum_{t=y}^{\infty} \frac{1}{N} \sum_{i=1}^{N} \frac{i}{N}(1-\frac{i}{N})^{t-1}$

    $= \frac{1}{N} \sum_{i=1}^{N} \frac{i}{N}(1-\frac{i}{N})^{y-1} \sum_{t=0}^{\infty} (1-\frac{i}{N})^{t}$

    $= \frac{1}{N} \sum_{i=1}^{N} \frac{i}{N}(1-\frac{i}{N})^{y-1} \frac{1}{1-(1-\frac{i}{N})}$

    $= \frac{1}{N} \sum_{i=1}^{N} (1-\frac{i}{N})^{y-1} \xrightarrow{\text{as } N\to\infty} \int_0^1 (1-x)^{y-1}dx$

    $= \frac{1}{y}$.

eg location-scale normal

- $x = (x_1, \ldots, x_n)' \overset{iid}{\sim} N(\mu, \sigma^2)$

$$\mu \mid \sigma^2 \sim N(\mu_0, \tau_0^2 \sigma^2)$$

$$\frac{1}{\sigma^2} \sim \text{gamma}_{\text{rate}}(\alpha_0, \beta_0) = \frac{1}{2\beta_0} \text{chi-squared}(2\alpha_0)$$

- mss $T(x) = (T_1(x), T_2(x)) = (\bar{x}, \|x - \bar{x}1\|^2)$

$\bar{x} \mid \mu, \sigma^2 \sim N(\mu, \frac{\sigma^2}{n})$ ind of $\|x - \bar{x}1\|^2 \mid \mu, \sigma^2 \sim \sigma^2 \text{chi-squared}(n-1) = \text{gamma}_{\text{rate}}(\frac{n-1}{2}, \frac{1}{2\sigma^2})$

- we check the prior on σ^2 first via

$$M_{T_2}(m_{T_2}(t_2) \le m_{T_2}(T_2(x)))$$

- so we need m_{T_2}. $\quad x = \mu 1 + \sigma z \quad \|x - \bar{x}1\|^2 = \sigma^2 \|z - \bar{z}1\|^2$

- $\|x - \bar{x}1\|^2 \sim X_1 / X_2$ where X_1, X_2 ind

$$\sim \frac{\text{chi-squared}(n-1)}{\text{gamma}_{\text{rate}}(\alpha_0, \beta_0)}$$

$$\frac{\beta_0^{\alpha_0}}{\Gamma(\alpha_0)} x_2^{\alpha_0 - 1} e^{-\beta_0 x_2} = \frac{1}{\Gamma(\frac{2\alpha_0}{2})} \left(\frac{2\beta_0 x_2}{2}\right)^{\frac{2\alpha_0}{2} - 1} e^{-\frac{2\beta_0 x_2}{2}} 2\beta_0$$

$$= \frac{1}{2\beta_0} \text{chi-squared}(2\alpha_0) \text{ density}$$

$\therefore \| y - \bar{x} 1 \|^2 \sim (n-1) \frac{2\beta_0}{2\alpha_0} \frac{X_1}{n-1} \Big/ \frac{2\beta_0 X_2}{2\alpha_0}$

$\sim (n-1) \frac{\beta_0}{\alpha_0} F(n-1, 2\alpha_0)$

$$\therefore m_{T_2}(t_2) = \frac{\Gamma\left(\frac{n-1+2\alpha_0}{2}\right)}{\Gamma\left(\frac{n-1}{2}\right)\Gamma\left(\frac{2\alpha_0}{2}\right)} \frac{\left(\frac{\alpha_0}{(n-1)\beta_0} \frac{n-1}{2\alpha_0} t_2\right)^{\frac{n-1}{2}-1}}{\left(1 + \frac{\alpha_0}{(n-1)\beta_0} \frac{n-1}{2\alpha_0} t_2\right)^{\frac{n-1+2\alpha_0}{2}}} \frac{1}{2\beta_0}$$

and then $M_{T_2}(m_{T_2}(t_2) \le m_{T_2}(T_2(x)))$

can be computed via simulation on t_2

- if the prior on σ^2 passes then check the prior on μ via

$M_{T_1}(m_{T_1}(t_1) \le m_{T_1}(T_1(x)))$

- so we need m_{T_1}

$T_1(x) = \bar{x} = \mu + \frac{\sigma}{\sqrt{n}} z = \mu_0 + \tau_0 \sigma u + \frac{\sigma}{\sqrt{n}} z$ (with $z \sim N(0,1)$, $u \sim N(0,1)$)

$= \mu_0 + \sigma\left(\tau_0 u + \frac{1}{\sqrt{n}} z\right)$

$\therefore \bar{x} \mid \sigma^2 \sim N\left(\mu_0, \sigma^2\left(\tau_0^2 + \frac{1}{n}\right)\right)$

$\therefore \frac{\bar{x} - \mu_0}{\sqrt{\tau_0^2 + 1/n}\,\sigma} = Z \Big/ \sqrt{X_2}$ (with $Z \sim N(0,1)$)

$$= \frac{N(0,1)}{\sqrt{\frac{2\alpha_0}{2\beta_0}\,\frac{\text{chi-squared}(2\alpha_0)}{2\alpha_0}}} = \sqrt{\frac{\beta_0}{\alpha_0}}\; t_{2\alpha_0}$$

$$\therefore\; m_{T_2}(t_2) = \frac{\Gamma\left(\frac{2\alpha_0+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{2\alpha_0}{2}\right)} \left(1 + \frac{\alpha_0}{\beta_0 2\alpha_0}\left(\frac{\bar{x}-\mu_0}{\sqrt{\tau_0^{-1}+1/n}}\right)^2\right)^{-\frac{2\alpha_0+1}{2}} \frac{1}{\sqrt{2\beta_0}\sqrt{\tau_0^{-1}+1/n}}$$

then compute $M_1\left(m_{T_1}(t_1) \le m_{T_1}(T_1(x))\right)$
via simulation from t_1

(4.) Inference

- different approaches to inference about $\psi = \Psi(\theta)$

(a) Bayesian Decision Theory

- we add another ingredient, namely a loss function $L : \Theta \times \Psi \to [0, \infty)$ where $L(\theta, a) = 0$ iff $a = \Psi(\theta)$

- but - this violates the principle of empirical criticism as there is no way to falsify the choice of L

  - we prefer theories of inference that don't contain unfalsifiable elements.

- in any case.

- the theory adds a principle

Principle of Minimized Expected Loss (or Maximized Expected Utility where utility fn is $U(\theta, a) = -L(\theta, a)$)

Choose decision fn $\delta : \mathcal{X} \to \Psi$ that minimizes

$$r(\delta) = E\left(L(\theta, \delta(x)) \right)$$

$$r(\delta) = \int_{\Theta}\int_{\mathcal{X}} L(\theta, \delta(x)) f_\theta(x)\,dx\,\pi(\theta)\,d\theta$$
$$= \int_{\Theta} R_\delta(\theta)\pi(\theta)\,d\theta.$$

where E is taken with respect to the joint distribution of (θ, x)

- $r(\delta)$ is called the <u>prior risk</u>
- δ_π satisfying $r(\delta_\pi) \leq r(\delta)$ $\forall \delta$ is called a <u>Bayes rule</u> and $r(\delta_\pi)$ is called the <u>Bayes risk</u>

note
<u>Lemma</u> A Bayes rule is obtained by minimizing the posterior risk for each $x \in \mathcal{X}$.

Proof. We have that

$$r(\delta) = \int_{\Theta}\int_{\mathcal{X}} L(\theta, \delta(x))\,\pi(\theta) f_\theta(x)\,dx\,d\theta$$

$$= \int_{\mathcal{X}}\int_{\Theta} L(\theta, \delta(x)) \frac{\pi(\theta) f_\theta(x)}{m(x)}\, m(x)\,d\theta\,dx$$

$$= \int_{\mathcal{X}} \left(\int_{\Theta} L(\theta, \delta(x))\,\pi(\theta|x)\,d\theta \right) m(x)\,dx$$

$$= \int_{\mathcal{X}} r(\delta|x)\, m(x)\,dx$$

and the result follows.

the posterior risk of δ.

eg <u>quadratic loss</u>

- suppose $\Psi \subseteq \mathbb{R}^k$ is convex and $L(\theta, a) = \|\Psi(\theta) - a\|^2$
- so we want $\delta(x) \in \Psi$ that minimizes

$$r(\delta|x) = E_{\pi(\cdot|x)}\left(\|\Psi(\theta) - \delta(x)\|^2\right)$$

$$= E_{\pi_\Psi(\cdot|x)}\left(\|\psi - \delta(x)\|^2\right)$$

$$= E_{\pi_\Psi(\cdot|x)}\left(\|\psi - E_{\pi_\Psi(\cdot|x)}(\psi) + E_{\pi_\Psi(\cdot|x)}(\psi) - \delta(x)\|^2\right)$$

$$= E_{\pi_\Psi(\cdot|x)}\left(\|\psi - E_{\pi_\Psi(\cdot|x)}(\psi)\|^2\right)$$

$$+ 2E_\sim\left((\psi - E_\sim(\psi))'(E_\Psi(\psi) - \delta(x))\right)$$

$$+ E_\sim\left(\|E_\sim(\psi) - \delta(x)\|^2\right)$$

$$= E_\sim\left(\| \quad \|^2\right) + 2\left(E_\sim(\psi) - E_\sim(\psi)\right)'\left(E_\sim(\psi) - \delta(x)\right) \quad (= 0)$$

$$+ \|E_\sim(\psi) - \delta(x)\|^2$$

$$\geq E\left(\|\psi - E_{\pi_\Psi(\cdot|x)}(\psi)\|^2\right) \text{ with equality}$$

iff $\delta(x) = E_{\pi_\Psi(\cdot|x)}(\psi)$

∴ provided the posterior expectation exists the Bayes rule is given by the posterior mean of ψ.

- but dependent on loss

eg hypothesis testing

- $\Theta = H_0 \cup H_a$, $H_0 \cap H_a = \phi$

- $L : \Theta \times \{0,1\} \longrightarrow \{0,1\}$ $\quad$ $\Psi(\theta) = I_{H_a}(\theta)$

where $L(\theta,\psi) = \begin{cases} 1 & \theta \in H_0, \psi = 1 \text{ or } \theta \in H_a, \psi = 0 \\ 0 & \text{otherwise} \end{cases}$

where $\delta(x) = 1$ means reject H_0

and $\delta(x) = 0$ means accept H_0

- $r(\delta|x) = E_{\pi(\cdot|x)}(L(\theta, \delta(x)))$

$$= \begin{cases} \pi(H_0 | x) & \text{if } \delta(x) = 1 \\ \pi(H_a | x) & \text{if } \delta(x) = 0 \end{cases}$$

∴ put $\delta(x) = 1$ $\quad \pi(H_0|x) \leq \pi(H_a|x)$
and 0 otherwise.

- so reject H_0 whenever posterior prob. of H_0 is ≤ posterior prob of H_a

- **but** - suppose $\pi(H_0) = 0$ say $H_0 = \Psi^{-1}\{\psi_0\}$
and π_Ψ is continuous then
$\pi(H_0) = \int_{\{\psi_0\}} \pi_\Psi(\psi)\, d\psi = 0$

- then $\pi(H_0 | x) = 0$ and we <u>always</u> reject H_0 which is ridiculous

- various fixes (later)

- <u>note</u> - no assessment of accuracy of estimate or strength of evidence

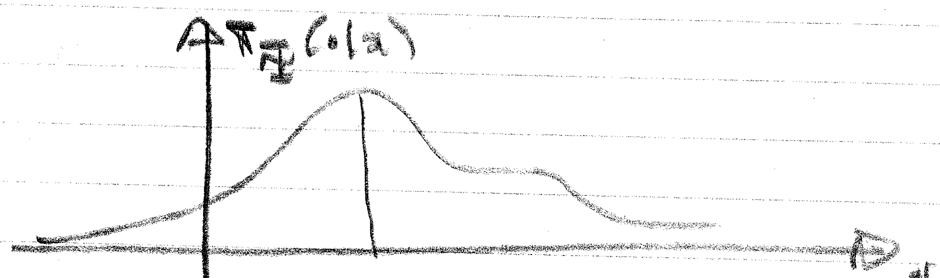
(b) **MAP (maximum a posteriori) inferences**

- inferences about $\psi = \Psi(\theta)$
- MAP preference ordering:

  we prefer ψ_1 to ψ_2 whenever $\pi_\Psi(\psi_1 | x) \geq \pi_\Psi(\psi_2 | x)$

- so we prefer ψ_1 to ψ_2 whenever our belief that ψ_1 is the true value is $\geq$ our belief that ψ_2 is the true value
- estimation: $\psi_{MAP}(x)$ satisfies

$$\pi_\Psi(\psi | x) \leq \pi_\Psi(\psi_{MAP}(x) | x), \quad \forall \psi \in \Psi$$

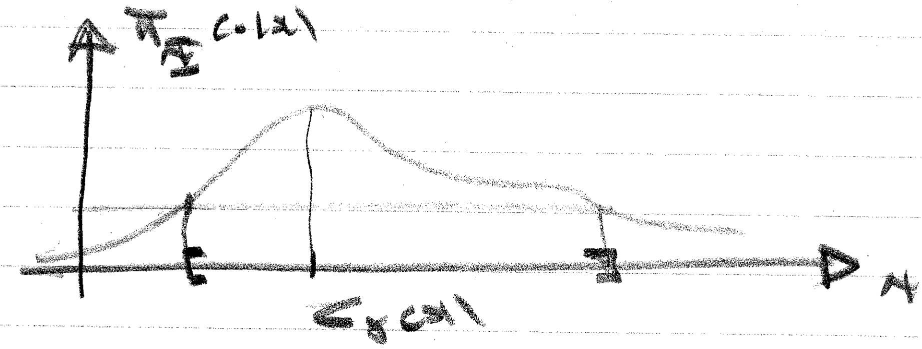


$\psi_{MAP}(x)$ = posterior mode

- to assess accuracy $\psi_{MAP}(x)$ compute the following γ-credible region where we choose $\gamma \in [0,1]$ (say $\gamma = 0.95$)

$$C_\gamma(x) = \{ \psi \mid \pi_\Psi(\psi | x) \geq c_\gamma(x) \}$$

- to see what $C_\gamma(x)$ is, let $G_\Psi(\cdot \mid x)$ be the cdf of $\pi_\Psi(\psi \mid x)$ when $\psi \sim \pi_\Psi(\cdot \mid x)$
- so $c_\gamma(x) = G_\Psi^{-1}(1-\gamma \mid x)$



- note $\psi_{MAP}(x) \in C_{\gamma_1}(x) \subseteq C_{\gamma_2}(x)$ whenever $\gamma_1 \le \gamma_2$
- so look at size of $C_\gamma(x)$ to assess the accuracy of $\psi_{MAP}(x)$

<u>eg</u> <u>location-scale normal</u>

- $x = (x_1, \ldots, x_n) \overset{iid}{\sim} N(\mu, \sigma^2)$
  $\mu \mid \sigma^2 \sim N(\mu_0, \tau_0^2 \sigma^2)$, $1/\sigma^2 \sim \text{gamma}_{\text{rate}}(\alpha_0, \beta_0)$
- we showed $\mu \mid x, \sigma^2 \sim N\left(\left(n+\frac{1}{\tau_0^2}\right)^{-1}\left(n\bar{x} + \frac{\mu_0}{\tau_0^2}\right), \left(n+\frac{1}{\tau_0^2}\right)\sigma^2\right)$

$$\frac{1}{\sigma^2} \Big| x \sim \text{gamma}_{\text{rate}}\left(\frac{n+2\alpha_0}{2}, \beta_0 + \frac{(n-1)s_x^2}{2} + \frac{1}{2}\frac{n}{n\tau_0^2+1}(\bar{x}-\mu_0)^2\right)$$

β_x

$\propto \chi^2_{n+2\alpha_0} / 2\beta_x$ density

$$\therefore \frac{\mu - \left(n + \frac{1}{\tau_0^2}\right)^{-1}\left(n\bar{x} + \frac{\mu_0}{\tau_0^2}\right)}{\sqrt{n + \frac{1}{\tau_0^2}}} \sim Z \Big/ \sqrt{\frac{1}{\sigma^2}}$$

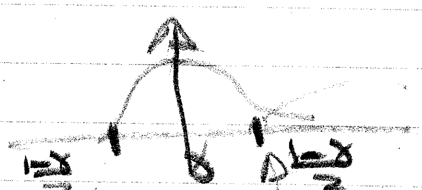
$$Z \Big/ \sqrt{(n+2\alpha_0)\, \frac{\chi^2_{n+2\alpha_0}}{2\beta_x (n+2\alpha_0)}}$$

$$= \sqrt{\frac{2\beta_x}{n+2\alpha_0}}\; t_{n+2\alpha_0}$$

- MAP estimator of μ is

$$\mu_{MAP}(x) = \left(n + \frac{1}{\tau_0^2}\right)^{-1}\left(n\bar{x} + \frac{\mu_0}{\tau_0^2}\right)$$

- γ-credible region is



$$\gamma + \frac{1-\gamma}{2} = \frac{1+\gamma}{2}$$

$$\left(n + \frac{1}{\tau_0^2}\right)^{-1}\left(n\bar{x} + \frac{\mu_0}{\tau_0^2}\right) \pm \sqrt{n + \frac{1}{\tau_0^2}}\,\sqrt{\frac{2\beta_x}{n+2\alpha_0}}\; t_{n+2\alpha_0,\, \frac{1+\gamma}{2}}$$

- <u>hypothesis assessment</u> $H_0 = \Psi^{-1}\{\psi_0\}$
- compute the Bayesian p-value

$$\Pi_\Psi\left(\pi_\Psi(\psi\,|\,x) \le \pi_\Psi(\psi_0\,|\,x)\,\middle|\,x\right)$$

and say evidence against H_0 when small

- <u>note</u> Bayesian p-value $< 1-\gamma$ iff $\psi_0 \notin C_\gamma(x)$

eg <u>location-scale normal</u> (cont'd)

- consider $\Psi(\mu, \sigma^2) = \mu$, $H_0: \mu = \mu_0$

$$\Pi_\Psi\left(\pi_\Psi(\mu\,|\,x) \le \pi_\Psi(\mu_0\,|\,x)\,\middle|\,x\right)$$

$$\pi_\Psi(\mu\,|\,x) = \frac{\Gamma\left(\frac{n+2\alpha_0+1}{2}\right)}{\Gamma\left(\frac{1}{2}\right)\Gamma\left(\frac{n+2\alpha_0}{2}\right)}\left(1 + \left(\frac{\mu - \overbrace{(n+\frac{1}{\tau_0^2})^{-1}(n\bar{x}+\frac{\mu_0}{\tau_0^2})}^{=\mu_x}}{\sqrt{\frac{2\beta_x}{n+2\alpha_0}}\sqrt{(n+1/\tau_0^2)^{-1}}}\right)^2\Big/(n+2\alpha_0)\right)^{-\frac{n+2\alpha_0+1}{2}}\frac{1}{\sqrt{2\beta_x}\sqrt{n+2\alpha_0}}$$

$$= \Pi_\Psi\left(\left(\frac{\mu-\mu_x}{\sqrt{\frac{2\beta_x}{n+2\alpha_0}}\sqrt{(n+1/\tau_0^2)^{-1}}}\right)^2 \ge \left(\frac{\mu_0-\mu_x}{\sqrt{\frac{2\beta_x}{n+2\alpha_0}}\sqrt{(n+1/\tau_0^2)^{-1}}}\right)^2\,\middle|\,x\right)$$

$$t \sim t_\alpha \quad t^2 \sim F(1, \alpha)$$

$$= 1 - F\left(1, n+2\alpha_0, \left(\frac{\mu_0-\mu_x}{\sqrt{\frac{2\beta_x}{\ldots}}\sqrt{\ldots}}\right)^2\right)$$

- problems with MAP

(1) - suppose $\lambda = \Lambda(\theta)$ where Λ is 1-1 and smooth

 - then posterior of λ is

$$\pi_{\Xi}(\Lambda^{-1}(\lambda) \mid x)\, J_{\Lambda}(\Lambda^{-1}(\lambda))$$

where $J_{\Lambda}(\theta) = \left(\det\left(\partial \Lambda(\theta)\right)\right)^{-1}$

- if $J_{\Lambda}(\theta)$ is not constant in θ then

$\lambda_{map}(x) \neq \Lambda(\theta_{map}(x))$ (generally)

and similarly for hpd regions

- we say MAP inferences are not <u>invariant</u> under reparametrization.

<u>eg</u> $x = (x_1, \ldots, x_n)' \sim$ Bernoulli(θ)

- $\theta \sim U(0,1)$

- $\pi(\theta) f_\theta(x) \propto \theta^{n\bar{x}} (1-\theta)^{n(1-\bar{x})}$

- $\therefore \theta \mid x \sim$ beta$(n\bar{x}+1, n(1-\bar{x})+1)$

$$\theta_{map}(x) = \frac{a-1}{a+b} = \frac{n\bar{x}}{n+2}$$

- $\lambda = \theta^2$, $\theta = \lambda^{1/2}$, $d\theta = \frac{1}{2}$

$$\lambda_{map}(x) = \left(\frac{n\bar{x} - 1}{n+1}\right)^2 \neq \theta^2_{map}(x)$$

(2) What is the evidence that H_0 is true?

- implicitly it is $\pi_{\Xi}(H_0 \mid x)$
- but this doesn't behave appropriately