ex] $n = 5$, $\sum_{i=1}^{5} \ln x_i = 9.6158$

x_i 0.5, 5, 10, 20, 30

Consider model Weibull ($\tau = 1.5$, $\theta = 25$)

vs Weibull ($\hat{\tau}, \hat{\theta}$),

The loglikelihood function of the Weibull fit at the optimum value is −17.8625.

perform LRT.

Soln] $f(x) = \dfrac{\tau x^{\tau-1}}{\theta^{\tau}} e^{-x^{\tau}/\theta^{\tau}}$

$$L(\tau,\theta) = \prod_{i=1}^{n} f(x_i) = \tau^{n} \prod_{i} x_i^{\tau-1} \theta^{-n\tau} e^{-\sum_i \frac{x_i^{\tau}}{\theta^{\tau}}}$$

$$\ln L(\tau,\theta) = n\ln(\tau) + (\tau-1)\sum_i \ln(x_i) - n\tau\ln(\theta) - \sum_i \frac{x_i^{\tau}}{\theta^{\tau}}$$

$\ln L(\tau = 1.5, \theta = 25) =$

$$5\ln(1.5) - (1.5-1)(9.6158) - 5(1.5)\ln(25) - \frac{\sum_i x_i^{1.5}}{(25)^{1.5}}$$

$= -19.6817$

i) H_0 Weibull ($\tau = 1.5$, $\theta = 25$) fit is good

H_1 use Weibull ($\hat{\tau}, \hat{\theta}$)

ii) $T = 2(\ln L_B - \ln L_A) = 2(-17.8625 + 19.6817) = 3.6384$

iii) $d = d_B - d_A = 2 - 0 = 2$

α	.1	.9	.05
d \ p	.9	.1	.95
2	.211	4.605	5.991

likely values: 4.605, 5.991

$T \leq 5.991$ (use $\alpha = .05$ when not given)

fail to reject Ho

notes 48.5

iv) use the Weibull($\tau = 1.5, \theta = 25$) model

§16.5

30]

$r =$ # of parameters	maximum log likelihood	Schwarz Bayesian (usually omitted)
1	$\ln(L_1)$	$\ln(L_1) - \frac{1}{2}\ln(n)$
2	$\ln(L_2)$	$\ln(L_2) - \frac{2}{2}\ln(n)$
$\vdots$	$\vdots$	
K	$\ln(L_K)$	$\ln(L_K) - \frac{K}{2}\ln(n)$.

31] Consider the table in 30]. The Schwarz Bayesian Criterion (SBC) says select the model that maximizes

$\ln(L_r) - \frac{r}{2}\ln(n)$ where n is the sample size and r is the number of parameters. So take the model that maximizes the 3rd column.

ex] $n=10$

$r=$ # parameters	$\ln(L_r)$	SBC $\ln(L_r)-\frac{r}{2}\ln(n)$
1	-321.32	$-321.32-1.1513=-322.4713$
2	-319.93	$-319.93-2.3026=-322.2326$ ←
3	-319.12	$-319.12-3.4539=-322.5739$
4	-318.12	$-318.12-4.605=-322.7252$

$0.5\ln(10)=1.1513$, The 2 parameter model is selected. See HW7 1b.

32] reject Ho if pvalue $<\alpha$
fail to reject Ho if pvalue $>\alpha$.
So reject Ho if pvalue is small.

§15.5 Bayesian Estimation

1] Frequentist methods assume $\underline{\theta}$ are constant parameters "generated by nature."
Bayesian Methods assume the parameters $\underline{\theta}$ are random variables.

ex] math 403 had $N|\lambda \sim \text{Poisson}(\lambda)$ and $\lambda \sim G(\alpha,\theta)$. Then $N \sim NB(\beta=\theta, r=\alpha)$.
This type of problem is closely related to Bayesian methods.

2) Let $\pi(\theta)$ be the prior pdf or pmf.

ex) $\pi(\lambda)$ is the $G(\alpha,\beta)$ pdf

λ: RV, dummy variable

α,β: "hyperparameters" $\approx$ usual parameters

Let $f(\underline{x}|\theta)$ be the conditional pdf or pmf (the usual likelihood function for frequentist models).

ex) $N|\lambda$ has the poisson(λ) pmf.

The joint pdf or pmf is

$f(\underline{x},\theta) = \pi(\theta)\, f(\underline{x}|\theta)$.

The posterior pdf or pmf is

$\pi(\theta|\underline{x}) = \dfrac{f(\underline{x},\theta)}{f(\underline{x})}$. For estimation, the prior and posterior are usually pdfs.

For testing the prior and posterior are usually pmfs.

$f(\underline{x}) = \int f(\underline{x},\theta)\, d\theta$ is the (marginal or) unconditional pdf or pmf.

ex] N has the NB($\beta=\theta$, $r=\alpha$) pmf. M404 50

3] <u>Bayes' Theorem</u>: The posterior distribution is given by

$$\pi(\underline{\theta}|\underline{x}) = \frac{f(\underline{x}|\underline{\theta})\,\pi(\underline{\theta})}{\int f(\underline{x}|\underline{t})\,\pi(\underline{t})\,d\underline{t}}$$

dummy variable, often written as $\underline{\theta}$

$$= \frac{f(\underline{x}|\underline{\theta})\,\pi(\underline{\theta})}{f(\underline{x})}.$$

Replace the integral by a sum for a pmf.

4] The posterior is just a conditional distribution: $\pi(\underline{\theta}|\underline{x}) = \dfrac{\text{joint of } \underline{x} \text{ and } \underline{\theta}}{\text{marginal of } \underline{x}}$.

Typically $f(\underline{x}|\underline{\theta}) = \prod_{i=1}^{n} f(x_i|\underline{\theta})$.

5] Often $\underline{\theta} = (\theta_1, \ldots, \theta_k)$ has $k=1$.

6] Often use $\pi(\theta|\underline{x}) \overset{\text{proportional to}}{\propto} f(\underline{x}|\theta)\,\pi(\theta) \propto$ brand name distribution, and then get the constant from the brand name distribution. Note $\theta \in$ interval for brand name distributions. Need a pdf to model a variable on an interval.

ex] Suppose that conditional on μ,
$x_1, \ldots, x_n \overset{ind}{\sim} \text{Exp}\left(\frac{1}{\mu}\right)$ and the

prior $\mu \sim G(\alpha, \frac{1}{\lambda})$.

So $\pi(\mu) = \dfrac{\mu^{\alpha-1} e^{-\mu\lambda}}{\left(\frac{1}{\lambda}\right)^{\alpha} \Gamma(\alpha)} \propto \mu^{\alpha-1} e^{-\mu\lambda}$.

$f(\underline{x}|\mu) = \prod_{i=1}^{n} f(x_i|\mu) = \prod_i \mu e^{-x_i\mu} = \mu^n e^{-\mu \sum_{i=1}^{n} x_i}$.

So $\pi(\mu|\underline{x}) \propto f(\underline{x}|\mu)\pi(\mu) \propto \mu^{n+\alpha-1} e^{-\mu\left(\lambda + \sum_{i=1}^{n} x_i\right)}$.

So the posterior pdf is the

gamma $\left(n+\alpha, \dfrac{1}{\lambda + \sum_{i=1}^{n} x_i}\right)$ pdf.

<u>warning</u>: You need to memorize some pdfs to do this quickly.

7] The prior distribution can be interpreted as the prior information about $\underline{\theta}$ before gathering data. After gathering data, the likelihood function is used to update the prior distribution, resulting in the posterior distribution for $\underline{\theta}$.

8] The posterior distribution is a perfectly good probability distribution. Let $W = \theta | \underline{x}$.

Then $P(a < W < b) = P(a < \theta < b | \underline{x})$

and $E(W) = E[\theta | \underline{x}]$.

9] If the prior distribution has support $[a,b]$, then the support of the posterior distribution is a subset of $[a,b]$ (often $[a,b]$).

ex] HW 7 4) The RV q models a probability, so $q \in [0,1]$. The prior and the posterior have support $(0.6, 0.8)$.

10] p.414 Often the posterior distribution is a brand name distribution. If a conjugate prior is used, then the posterior distribution has the same distribution as the conjugate prior, but different parameters. Conjugate priors tend to go with exponential families.

ex] Often θ models a probability.

So $\theta \in [0,1]$.

$W \sim \text{beta}(a,b)$ if $f(w) = \frac{\Gamma(a+b)}{\Gamma(a)\Gamma(b)} w^{a-1}(1-w)^{b-1}$

where $a > 0$, $b > 0$, $w \in [0,1]$.

For the geometric and negative binomial, can take $\rho = \frac{1}{1+\beta} \in [0,1]$ and the binomial has $\theta \in [0,1]$.

If x_i are conditionally iid and $P(x_i = 1) = q = 1 - P(x_i = 0)$, then the x_i are Bernoulli (q) = binomial $(q, m=1)$.

If x_i are conditionally iid $\text{bin}(q,m)$, $f(\underline{x}|q) \propto q^{\sum_{i=1}^n x_i}(1-q)^{nm - \sum_{i=1}^n x_i}$.

If the prior $\pi(q) \sim \text{beta}(a,b)$, then the posterior

$$\pi(q|\underline{x}) \propto \pi(q) f(\underline{x}|q) \propto q^{a+\sum_i x_i - 1}(1-q)^{b+nm-\sum_i x_i - 1}.$$

So $q|\underline{x} \sim \text{beta}\left(a + \sum_i x_i,\ b + nm - \sum_i x_i\right)$.

If $X \sim \text{bin}(q,m)$ $(n=1)$,

then $q|X \sim \text{beta}(a+X,\ b+m-X)$.

If $m=1$ (Bernoulli trials), then

$q|\underline{x} \sim \text{beta}\left(a+\sum_i x_i, b+n-\sum_i x_i\right)$.

11) Notation: text is often using capital Greek letters. So Θ is the RV and $\Theta = \theta$ is the observed RV. Some Greek letters are ugly like $\Xi = \xi$.

$\underbrace{}_{X_i}$

12)* a) $\theta|\underline{x} \sim \text{beta}(a,b)$ if $\pi(\theta|\underline{x}) \propto \theta^{a-1}(1-\theta)^{b-1}$ and $\theta \in [0,1]$, $a,b>0$

b) $\theta|\underline{x} \sim \text{gamma}(\alpha,\beta)$ if $\pi(\theta|\underline{x}) \propto \theta^{\alpha-1} e^{-\theta/\beta}$, $\alpha,\beta,\theta>0$.
$\text{Exp}(\beta)$ has $\alpha=1$.

c) $\theta|\underline{x} \sim$ single parameter pareto(α,β) if
$\pi(\theta|\underline{x}) \propto \theta^{-(\alpha+1)}$ where $\theta > \beta$, $\alpha>0$, $\beta \in \mathbb{R}$.

d) $\theta|\underline{x} \sim$ pareto(α,β) if

$\pi(\theta|\underline{x}) \propto (\beta+\theta)^{-(\alpha+1)}$, $\alpha,\beta,\theta>0$.

e) $\theta|\underline{x} \sim N(\mu,\sigma^2)$ if

$\pi(\theta|\underline{x}) \propto \exp\left(\frac{-1}{2\sigma^2}(\theta-\mu)^2\right)$, $\sigma^2>0$, $\theta,\mu \in \mathbb{R}$.

Note θ takes the place of x and β often takes the place of θ in the pdf.

13] Recognizing that $\pi(\theta|\underline{x}) \propto \pi(\theta) f(\underline{x}|\theta)$ is a brand name distribution is easier than calculating the constant c that makes $\pi(\theta|\underline{x})$ a pdf (or pmf).

ex] Losses follow a $U(0,\theta)$ dist.

$$\pi(\theta) = \frac{24}{\theta^4}, \quad \theta \geq 2.$$

(parameter other than θ would be nice.)

a) Find the posterior distribution after 10 losses with a maximum loss of 5.

So(n) $f(\underline{x}|\theta) \propto \frac{1}{\theta^n} I[\theta \geq x_{(n)}]$.

So $\pi(\theta|\underline{x}) \propto \pi(\theta) f(\underline{x}|\theta) \propto \frac{1}{\theta^{n+4}} I[\theta \geq \max(x_{(n)}, 2)]$

$I(\theta > a) I(\theta > b) = I[\theta > \max(a,b)]$

$$\underbrace{= \frac{1}{\theta^{14}} I[\theta \geq 5]}_{\propto \text{ single parameter pareto}(\alpha = 13,\ \beta = 5) \sim W}$$

b) Find the mean of the above posterior dist.

So(n) $E(W) = \frac{\alpha\beta}{\alpha - 1} = \frac{13(5)}{12} = \boxed{\frac{65}{12}} = E[\theta|\underline{x}]$.