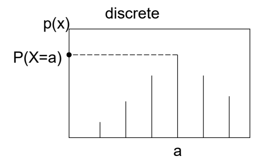
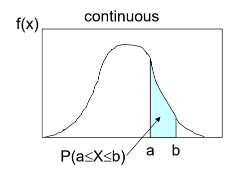
Continuous Bayes Methods

Decision Theory

Some differences between discrete/continuous

- Probability is not height of function but area under curve for an interval.

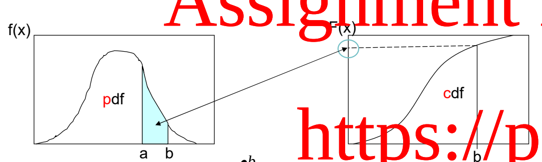




- Require f(X)≥0 for all X (but X can be < 0)
- Total area under the function =1

Probability vs. Cumulative Probability

PDF's and CDF's



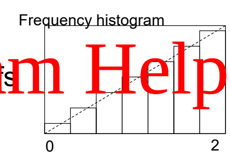
$$PDF: P(a \le X \le b) = \int_a^b f(X)dX$$

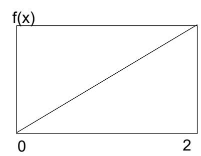
$$CDF: P(X \le x) = \int_{-\infty}^{x} f(\tau)d\tau$$

$$f(X) = \frac{d}{dX}F(X)$$

- P(a≤X≤b) = P(X≤b) – P(X≤a)
- P(X≥a) = 1-P(X≤a)
- P(X≤a) = 1-P(X≥a)

Empirical pdf's





$$f(X) = kX \qquad 0 \le X \le 2$$

$$\int_0^2 kXdX = k\frac{X^2}{2} = \frac{kX^2}{2}\Bigg|_0^2 = 2k$$

$$2k \equiv 1 \text{ so } k = \frac{1}{2}$$

$$\therefore f(X) = \frac{1}{2}X \qquad 0 \le X \le 2$$

Empirical pdfs

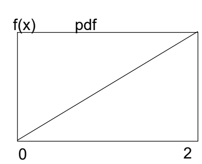
$$\mu = E[X] = \int_0^2 X \cdot \frac{1}{2}XdX = \frac{X^3}{6}\Bigg|_0^2 = \frac{4}{3}$$

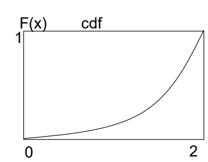
$$\sigma^2 = E[X^2] - \mu^2 = \int_0^2 X^2 \cdot \frac{1}{2}XdX - \mu^2$$

$$= \frac{X^4}{8}\Bigg|_0^2 - \left(\frac{4}{3}\right)^2 = 2 - \left(\frac{16}{9}\right) = \frac{2}{9}$$

Empirical pdfs: Cumulative Distribution Function (CDF)

$$F(X) = \int_0^x \frac{1}{2}\tau d\tau = \frac{\tau^2}{4}\Bigg|_0^x = \frac{1}{4}X^2 \qquad 0 \le X \le 2$$





Empirical pdfs: Expected value of a function
Y=Cost(X)=aX+b

$$E(Cost(X)) = \int_0^2 Cost(X)\frac{1}{2}XdX$$

$$= \int_0^2 (aX+b)\frac{1}{2}XdX = \int_0^2 \left(\frac{aX^2}{2}+\frac{b}{2}X\right)dX$$

$$= \left(\frac{aX^3}{6}+\frac{bX^2}{4}\right)\Bigg|_0^2 = \frac{4}{3}a+b$$

Or use properties of expectation:
Y=Cost(X)=aX+b

$$E(Cost(X)) = E(aX+b)$$
$$= E(aX)+E(b)$$
$$= aE(X)+b$$
$$= \frac{4}{3}a+b$$

Pdf's with more than one random variable

- f(X,Y) called the joint probability of X and Y
- Marginal pdfs can be obtained by integrating one variable out of the function:

$$f(X) = \int_{-\infty}^{\infty} f(X,Y)dY$$

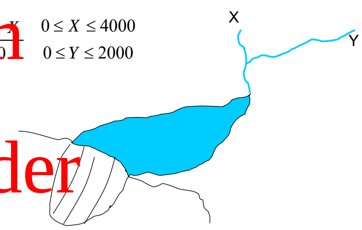
$$f(Y) = \int_{-\infty}^{\infty} f(X,Y)dX$$

Example

Suppose we have a reservoir fed by 2 rivers. Let X=stream flow in river X and Y=flow in river Y. The joint pdf of flows is:

$$f(X,Y) = c\frac{4000-X}{4000} \quad \begin{array}{l} 0 \le X \le 4000 \\ 0 \le Y \le 2000 \end{array}$$

$$c = 2.5x10^{-7}$$



Example

$$f(X) = \int_0^{2000} c\frac{4000-X}{4000}dY = \frac{c}{2}(4000-X)$$

$$f(Y) = \int_0^{4000} c\frac{4000-X}{4000}dX = 2000c$$

$$E[(X+Y)] = \int_0^{2000}\int_0^{4000} (X+Y)\cdot c\frac{4000-X}{4000}dXdY$$

Conditional Probability

$$f(X|Y) = \frac{f(X,Y)}{f(Y)}$$

Bayes rule is derived from the above conditional probability relationship where Y represents the sample information (observations) and X is the random (decision) variable.

ISE 562, Dr. Smith

Bayes for Continuous Random Variables

i) Conditional Probability $f(\theta \mid y) = \dfrac{f(\theta, y)}{f(y)}$

ii) As in discrete case: $f(\theta, y) = f(\theta) f(y \mid \theta)$

iii) Integrating out the θ $f(y) = \int_{-\infty}^{\infty} f(\theta, y) d\theta$

$$= \int_{-\infty}^{\infty} f(\theta) f(y \mid \theta) d\theta$$

Substituting (ii) and (iii) in (i)

$$f(\theta \mid y) = \frac{f(\theta) f(y \mid \theta)}{\int_{-\infty}^{\infty} f(\theta) f(y \mid \theta) d\theta}$$

9/4/2022 13

ISE 562, Dr. Smith

$$f(\theta \mid y) = \frac{f(\theta) f(y \mid \theta)}{\int_{-\infty}^{\infty} f(\theta) f(y \mid \theta) d\theta}$$

$$posterior\ pdf = \frac{(prior\ pdf)(likelihood)}{\int (prior\ pdf)(likelihood)}$$

9/4/2022 14

ISE 562, Dr. Smith

Sufficiency

- We denote an uncertain decision variable as θ and assume that sample information involving θ can be summarized by a sample statistic y.
- If y has all the information from the sample relevant to the uncertainty about θ, then y is called a *sufficient* statistic.
- Example: for a Bernoulli process, the sample information can be summarized by n and r; the actual number of successes and failures adds no additional information about p since p=r/n.

9/4/2022 15

ISE 562, Dr. Smith

Sufficiency

Why do we care about sufficiency?

For Bayes rule this means that knowledge of n and r is sufficient to determine the likelihoods so the posterior distribution of p given n and r is exactly the same as the posterior distribution of p given the entire sequence of observations.

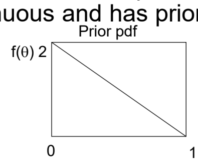
Simply put, everything we need to know about the sample is contained in the likelihood function.

9/4/2022 16

ISE 562, Dr. Smith

- Example: Let θ=market share of a new product (0≤θ≤1); assume it is continuous and has prior pdf:

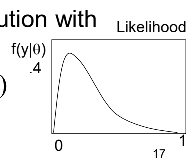
$$f(\theta) = 2(1-\theta) \qquad 0 \le \theta \le 1$$



- We take a sample of 5 consumers and 1 buys new brand while other 4 purchase a different brand
- Assume binomial likelihood distribution with success = “buys new product” so

$$f(y \mid \theta) = P(r = 1 \mid n = 5, \hat{\theta} = \theta)$$

$$= \frac{5!}{4! \cdot 1!} \theta^1 (1-\theta)^4 = 5\theta(1-\theta)^4$$



9/4/2022 17

ISE 562, Dr. Smith

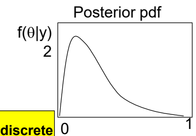
- Applying Bayes rule we substitute the prior and likelhood functions:

$$f(\theta \mid y) = \frac{f(\theta) f(y \mid \theta)}{\int_{-\infty}^{\infty} f(\theta) f(y \mid \theta) d\theta} = \frac{[2(1-\theta)][5\theta(1-\theta)^4]}{\int_0^1 [2(1-\theta)][5\theta(1-\theta)^4] d\theta}$$

$$= \frac{10\theta(1-\theta)^5}{10\int_0^1 \theta(1-\theta)^5 d\theta} = \frac{\theta(1-\theta)^5}{\int_0^1 \theta(1-\theta)^5 d\theta}$$

$$\left(Since \int_0^1 x^{m-1}(1-x)^{n-1} dx = \frac{\Gamma(m)\Gamma(n)}{\Gamma(m+n)}\ where\ \Gamma(n) = (n-1)! \right)$$

$$= \frac{\theta(1-\theta)^5}{\frac{1!5!}{7!}} = 42\theta(1-\theta)^5$$



Notice this is a function of θ. In the discrete Bayes case we had specific values of p for the states—here we can have <u>any</u> value in [0,1] for the state variable.

9/4/2022 18

- As shown the process appears somewhat difficult and computationally challenging
- May not find the integral is computable in closed form; may need numerical quadrature techniques

Example

- There is another way!!!
- $\Rightarrow$ conjugate families of distributions that simplify the process of combining priors and likelihoods

Properties of conjugate families of pdfs

1. Tractability: easy to specify posterior given the prior and likelihood function
2. Richness: the prior should reflect the prior information (this is done with parameters to fit the distribution to the information)
3. Ease of interpretation: prior should be interpretable in terms of previous sample results

2 conjugate families studied here

1. Sampling from a Bernoulli process (Likelihood) whose conjugate is the family of beta distributions (Prior)
2. Sampling from a normal distributed (Likelihood) process with known variance whose conjugate distribution is the family of normal distributions. (Prior)

"Conjugate" family refers to the relationship between the prior and the likelihood function.

1. Sampling from a Bernoulli process (the likelihood function) has as its conjugate, the family of beta distributions

$$f(p)=\frac{(n-1)!}{(r-1)!(n-r-1)!}p^{r-1}(1-p)^{n-r-1} \quad 0\le p\le 1$$

$n>r>0$

(Note that the random variable p, varies from zero to one for the beta pdf)

Beta pdf

"If n, r, not integers we must use gamma functions

$$f(p)=\frac{\Gamma(n)}{\Gamma(r)\Gamma(n-r)}p^{r-1}(1-p)^{n-r-1} \quad 0\le p\le 1$$

$$where$$

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

$$note: if\ t\ an\ integer$$

$$\Gamma(t)=(t-1)!$$

Mean and variance of the beta distribution:

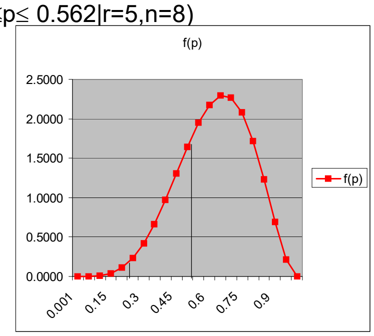
$$\mu=E(\hat{p}\,|\,r,n)=\frac{r}{n}$$

$$\sigma^2=V(\hat{p}\,|\,r,n)=\frac{r(n-r)}{n^2(n+1)}$$

To calculate probabilities we use fractiles:

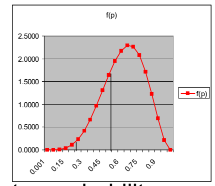
The f fractile of the pdf of a continuous random variable is the value x_f where $P(X\le x_f)=f$

Suppose we have a beta pdf of p with r=5 and n=8 and we want P(p≤0.562|r=5,n=8) and P(.265≤p≤ 0.562|r=5,n=8)



Suppose we have a beta pdf of p with r=5 and n=8

- Can use rhs of beta calculator (back of book) P(p≤0.562|r=5,n=8)=0.3402
- For P(.265≤p≤ 0.562|r=5,n=8) use:
- P(a ≤x≤ b)=P(x≤b) - P(x≤a)
- = P(p≤.562) - P(p≤.265)
- =0.3402 – 0.0167
- =0.3235



- Can use fractiles (on lhs) to enter probability and then look up p.

Fractile Calculator

Notatation Alert!

- Prior pdfs and likelihoods denoted with single primes; e.g., (p’), f’(θ), E’[θ], μ’,…
- Posterior pdfs and parameters of posterior pdfs denoted with double primes, (p”), f”(θ), E”[θ], μ”,…

Now the big advantage of conjugate pdfs…

Given a binomial process

- **Stationarity and Independency**
- **Beta prior pdf of the form:**

$$f'(p)=\frac{(n'-1)!}{(r'-1)!(n'-r'-1)!}p^{r'-1}(1-p)^{n'-r'-1} \quad 0\le p\le 1$$

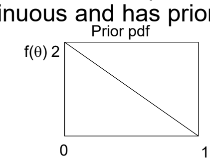
- **We draw a sample for binomial likelihood function with r successes in n trials. Then the posterior pdf becomes:**

$$f''(p)=\frac{(n''-1)!}{(r''-1)!(n''-r''-1)!}p^{r''-1}(1-p)^{n''-r''-1} \quad 0\le p\le 1$$

with n" = n' + n and r" = r' + r

- Example: Let θ=market share of a new product (0≤θ≤1); assume it is continuous and has prior pdf:

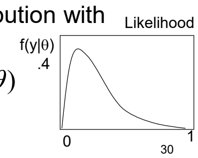
$$f(\theta)=2(1-\theta) \qquad 0\le\theta\le 1$$



- We take a sample of 5 consumers and 1 buys new brand while other 4 purchase a different brand
- Assume binomial likelihood distribution with success = “buys new product” so

$$f(y\mid\theta)=P(r=1\mid n=5,\hat{\theta}=\theta)$$

$$=\frac{5!}{4!\cdot 1!}\theta^1(1-\theta)^4=5\theta(1-\theta)^4$$

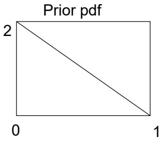


ISE 562, Dr. Smith — Beta/binomial

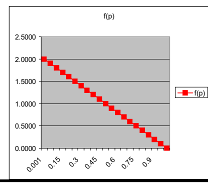
- Example: Let θ=market share of a new product (0≤θ≤1) is a beta prior:

$$f(\theta)=2(1-\theta) \qquad 0\le\theta\le 1$$

$$E[\theta]=\int_0^1 \theta 2(1-\theta)d\theta=\frac{1}{3}$$



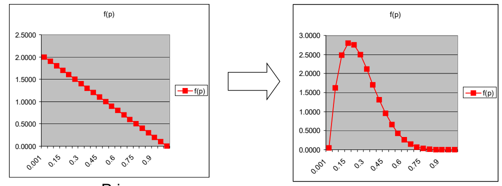
- The equivalent mean for the beta is E[p]=r/n=1/3 so r=1 and n=3:



9/4/2022 31

ISE 562, Dr. Smith — Beta/binomial

- The sample results were r=1 and n=5 so the posterior pdf has parameters
- r"=r'+r= 1+1=2
- n"=n'+n=3+5=8
- So mean of prior went from 1/3 (0.33) to posterior of 2/8=1/4 (0.25) – it shifted to the left.

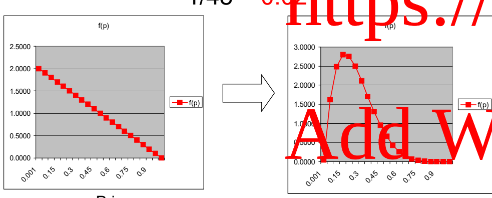


Prior Posterior

9/4/2022 32

ISE 562, Dr. Smith — Beta/binomial

- Notice also the shift in variance:
- Prior variance =r(n-r)/(n²(n+1))=1(3-1)/(9(4)) =1/18 = 0.055
- Posterior variance=2(8-2)/(64(9))=12/576 =1/48 = 0.02

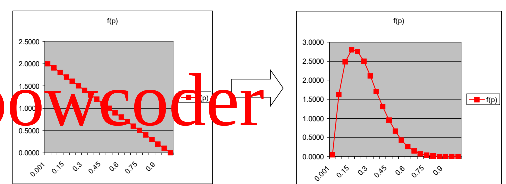


Prior Posterior

9/4/2022 33

ISE 562, Dr. Smith — Beta/binomial

- Notes:
- Posterior mean always lies between prior mean and the sample mean for the Bernoulli/Beta family
- Posterior variance generally smaller than prior variance due to addition of sample information to prior knowledge.



Prior Posterior

9/4/2022 34

ISE 562, Dr. Smith — Normal Conjugate

- “Conjugate” family refers to the relationship between the prior and the likelihood function.
- Sampling from a normal process (the likelihood function) has as its conjugate, the family of normal distributions

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

$$z=\frac{x-\mu}{\sigma} \qquad f(x\mid\mu,\sigma^2)=\frac{f(z\mid 0,1)}{\sigma}$$

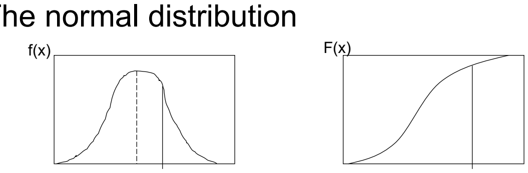
$$f(z)=\frac{1}{\sqrt{2\pi}}e^{-\frac{z^2}{2}}$$

Back to p. 54

9/4/2022 35

ISE 562, Dr. Smith

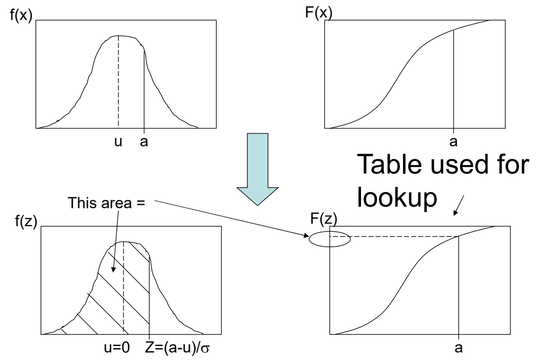
The normal distribution



$$P(X\le a)=\int_0^a f(X)dX$$

$$P(z\le\frac{(a-\mu)}{\sigma})=STDNORMALTABLE(\frac{(a-\mu)}{\sigma})$$

9/4/2022 36



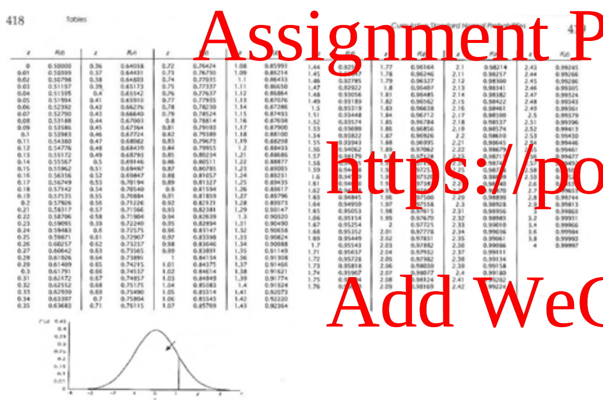
Depending on form of table, can use symmetry properties of normal to calculate various probabilities:

$P(a \le X \le b) = P(X \le b) - P(X \le a)$

$P(X \ge a) = 1 - P(X \le a)$

$P(X \le a) = 1 - P(X \ge a)$

Standard Normal Table



Two types of normal pdf problems:

1. Given X, u, σ^2, find P(X)
   a) Transform question to probability statement
   b) Translate X to Z and P(Z) using $Z=(X-u)/\sigma$
   c) Look up P(Z) in standard normal table
2. Given P(X), find X, u, or σ^2
   a) Lookup P(Z) in body of table to find Z
   b) Solve $Z=(X-u)/\sigma$ for unknown quantity

Type 1, Find Probability

A carpet warehouse keeps 6000 yards of carpet in stock during a month. The average demand is normally distributed with mean 4500 yards and standard deviation 900 yards. What is the probability a customer order won't be met? We want $P(d \ge 6000)$.

$P(d \ge 6000) = P(z \ge (6000-4500)/900) = P(z \ge 1.67)$

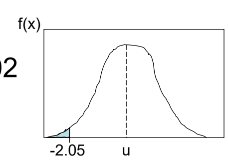
$= 1 - P(z \le 1.67) = 1 - (0.9525) = \underline{0.0475}$

Std. Normal Table

Type 2, Find unknown

Amount of coffee a filling machine puts into 4 oz. jars is normally distributed with std. dev, $\sigma=0.04$ oz. If only 2% of the jars are to contain less than 4 oz., what should be the average fill amount?

- We want $P(X \le 4) = 0.02$.
- Find Z such that $P(Z \le z) = 0.02$
- From table, Z=-2.05
- Solve $-2.05=(4-u)/.04$ for u.
- u=4.082 oz.



Std. Normal Table

Back to Bayes…

- μ and σ are summary measures of the normal pdf
- For the binomial, r and n are summary measures of the sample info
- For the normal, the sample mean and sample variance are summary measures for a sample from a normal population

$$m = \frac{\sum_{i=1}^{n} x_i}{n}$$

$$s^2 = \frac{\sum_{i=1}^{n} (x_i - m)^2}{n-1}$$

- If $x_1, \ldots, x_n$ random variables represent a random sample from a normal population with mean μ and σ, then the sample mean is normally distributed with $E[m|\ \mu,\sigma^2] = \mu$ and $V[m|\ \mu,\sigma^2] = \sigma^2/n$
- Even if population is not normally distributed the Central Limit Theorem shows that as $n \to \infty$, the distribution of $(m-\mu)/(\sigma/n)$ converges to the normal pdf. This is true for any population with finite mean and variance.

CLT demo

- Be careful—check your data for normal distribution

Chi-squared χ^2 Goodness of fit test

- Probability distribution unknown
- Sample from the unknown pdf n values
- We sort the sample into k class intervals (histogram) and let f_{oi} be the observed frequency (count) for interval i
- Then compute the theoretical frequency using normal (or any) distribution for each interval, f_{ti}
- If any of intervals contain < 5 theoretical observations, combine with next interval
- Calculate total deviation of observed values from theoretical values and test for significance

The Chi-square statistic is:

$$\chi^2 = \sum_{i=1}^{k} \frac{(f_{oi} - f_{ti})^2}{f_{ti}}$$

k= no. intervals
p= no. estimated parameters (2 for normal)

The hypothesis test:

Ho: X is normally distributed

Ha: X is not normally distributed

Reject Ho if data not normal (if differences too large); ie: $\textit{reject normal if } \chi^2 > \chi^2_{\alpha,k-p-1}$

Example: cell phone company has frequency data for length of calls outside a roaming area. The mean and std. dev are 14.3 and ±3.7 minutes respectively. Are the data normally distributed? Use α=0.05.

Length (min)	Frequency
0-5	26
5-10	75
10-15	139
15-20	105
20-25	37
25+	18
Total	400

Testing for Normality

P(0≤X≤5)=P(z≤(5-14.3)/3.7) - P(z≤(0-14.3)/3.7)=P(z≤-2.51)-P(z≤-3.86)=(1-.9940)-(1-.9999)=0.006

Length (min)	Freq (obs)	Theor. prob	Theor freq (np)	$(f_o-f_t)^2/f_t$
0-5	26	.006	2.4	
5-10	75 } 101	.117	46.8 } 49.2	54.5
10-15	139	.452	180.8	9.7
15-20	105	.363	145.2	11.2
20-25	37	.06	24 } 24.8	36.8
25+	18 } 55	.002	.8	
Total	400			112.2

Testing for Normality

$$\text{reject normal if } \chi^2 > \chi^2_{\alpha,k-p-1}$$

$$\chi^2 = 112.2$$

$$\chi^2_{\alpha,k-p-1} = \chi^2_{0.05,4-2-1} = \chi^2_{0.05,1} = 3.841$$

$$112.2 > 3.841$$

$$\text{so reject normal}$$

Example

Normal Conjugate

Returning to conjugate discussion…

Suppose the prior distribution on μ is normal with mean m' and variance σ'²:

$$f'(\mu) = \frac{1}{\sqrt{2\pi\sigma'^2}} e^{-\left(\frac{(\mu-m')^2}{2\sigma'^2}\right)}$$

We draw a sample of size n and observe a sample mean of m; the posterior density is then normal:

$$f''(\mu \mid y) = \frac{1}{\sqrt{2\pi\sigma''^2}} e^{-\left(\frac{(\mu-m'')^2}{2\sigma''^2}\right)}$$

Where y represents the sample results and the posterior parameters are computed from:

$$\frac{1}{\sigma''^2} = \frac{1}{\sigma'^2} + \frac{n}{\sigma^2} \quad and \quad m'' = \frac{\frac{m'}{\sigma'^2} + \frac{n}{\sigma^2}m}{\frac{1}{\sigma'^2} + \frac{n}{\sigma^2}}$$

Back to p. 56

Normal Conjugate

Discrete prior example: retailer interested in weekly sales, x at one of their stores

- x distributed normally with unknown mean μ and variance known, σ²=90000
- Only 5 potential values to be considered
- μ=1100, 1150, 1200, 1250, 1300
- The prior distribution is estimated to be:

P(μ=1100)=.15
P(μ=1150)=.20
P(μ=1200)=.30
P(μ=1250)=.20
P(μ=1300)=.15

Normal Conjugate

Retailer wants more information about the store so takes a sample from past sales records assuming weekly sales independent.

Takes sample of n=60 weeks and calculates sample mean, m=1240; now calculate the likelihoods*

$$f(1240 \mid \mu=1100, \frac{\sigma}{\sqrt{n}}=38.73) = f\left(\frac{1240-1100}{38.73} \mid 0,1\right)/38.73 = .0006/38.73$$

$$f(1240 \mid \mu=1150, \frac{\sigma}{\sqrt{n}}=38.73) = f\left(\frac{1240-1150}{38.73} \mid 0,1\right)/38.73 = .0270/38.73$$

$$f(1240 \mid \mu=1200, \frac{\sigma}{\sqrt{n}}=38.73) = f\left(\frac{1240-1200}{38.73} \mid 0,1\right)/38.73 = .2347/38.73$$

$$f(1240 \mid \mu=1250, \frac{\sigma}{\sqrt{n}}=38.73) = f\left(\frac{1240-1250}{38.73} \mid 0,1\right)/38.73 = .3857/38.73$$

$$f(1240 \mid \mu=1300, \frac{\sigma}{\sqrt{n}}=38.73) = f\left(\frac{1240-1300}{38.73} \mid 0,1\right)/38.73 = .1200/38.73$$

* 300/sqrt(60) = 38.73

Normal Conjugate

Note that denominator same for all calculations (see page 35) so we can ignore in table—its optional since it cancels out later

μ	Prior prob.	Likelihood	Prior prob * likelihood	Posterior prob.
1100	.15	.0006	.00009	.001
1150	.20	.0270	.00540	.032
1200	.30	.2347	.07041	.412
1250	.20	.3857	.07714	.450
1300	.15	.1200	.01800	.105
	1.00		.17104	1.000

ISE 562, Dr. Smith — Normal Conjugate

Continuous prior example: Suppose mgr decides prior is normally distributed with mean, m'=1200 and σ'=50.

Note that σ^2=90000 is the variance of weekly sales

Note that σ'^2=2500 is the variance of prior pdf of μ, *average* weekly sales.

Using the previous sample information for n=60 weeks with mean m=1240 we calculate the posterior parameters for the posterior pdf:

9/4/2022 55

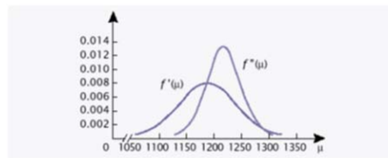
ISE 562, Dr. Smith — Normal Conjugate

$$\frac{1}{\sigma''^2}=\frac{1}{\sigma'^2}+\frac{n}{\sigma^2}=\frac{1}{2500}+\frac{60}{90000}=\frac{96}{90000}$$

Ref p. 51

$$m''=\frac{\frac{1}{\sigma'^2}m'+\frac{n}{\sigma^2}m}{\frac{1}{\sigma'^2}+\frac{n}{\sigma^2}}=\frac{\frac{1}{2500}1200+\frac{60}{90000}1240}{\frac{1}{2500}+\frac{60}{90000}}=1225$$

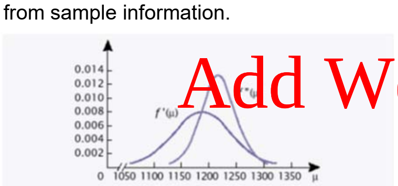
So the mean and variance of the posterior pdf are 1225 and 90000/96=937.5



9/4/2022 56

ISE 562, Dr. Smith — Normal Conjugate

- The original belief about the mean (prior) was m'=1200 and a population standard deviation of 300
- A sample was taken with sample mean of 1240 and standard deviation of 38.73
- The posterior mean then moved from the prior value of 1200 to the posterior value of 1225
- The posterior standard deviation went from a prior value of 38.73 to sqrt(937.5) = 30.62 (less dispersion due to "learning" from sample information.



9/4/2022 57

ISE 562, Dr. Smith — What if no conjugates?

- What if the prior and likelihood not conjugate?
- All is not lost. In Bayes time until digital computers arrived in the 1950-1960's, computation of the posterior for an arbitrary prior and likelihood was not feasible.
- Today we have software that make these calculations easy (e.g., Mathematica, Matlab, etc.)

Example

9/4/2022 58