Meeting 3:
Exponential class of distributions,
Interpretation of priors

The exponential class of distributions

A (family) of probability distribution(s) belong(s) to the k-parameter exponential class of distributions if the probability density (or mass) function can be written:

$$f(\boldsymbol{x}|\boldsymbol{\theta}) = e^{\sum_{j=1}^{k} A_j(\boldsymbol{\theta})B_j(\boldsymbol{x})+C(\boldsymbol{x})+D(\boldsymbol{\theta})}$$

where

- $\boldsymbol{\theta} = (\theta_1 , \dots , \theta_k)$
- $A_1(\boldsymbol{\theta}), \dots, A_k(\boldsymbol{\theta})$ and $D(\boldsymbol{\theta})$ are functions of the parameter $\boldsymbol{\theta}$ only (and not of $\boldsymbol{x}$)
- $B_1(\boldsymbol{x}), \dots , B_k(\boldsymbol{x})$ and $C(\boldsymbol{x})$ are functions of $\boldsymbol{x}$ only (and not of $\boldsymbol{\theta}$)

Examples

Two parameter Gamma distribution (univariate), shape and rate parameterization:

$$f(x|\theta) = f(x|\alpha,\beta) = \frac{\beta^{\alpha}}{\Gamma(\alpha)} x^{\alpha-1} e^{-\beta x} \qquad ; \; x \geq 0$$

$$= e^{(\alpha-1)(\ln x)-\beta x+\alpha \ln \beta-\ln \Gamma(\alpha)} = e^{\alpha \ln x-\beta x-\ln x+\alpha \ln \beta-\ln \Gamma(\alpha)}$$

Parametric form 1: $e^{(\alpha-1)(\ln x)-\beta x+\alpha \ln \beta-\ln \Gamma(\alpha)}$

$$A_1(\boldsymbol{\theta}) = A_1(\alpha,\beta) = \alpha - 1$$
$$A_2(\boldsymbol{\theta}) = A_2(\alpha,\beta) = \beta$$
$$B_1(\boldsymbol{x}) = \ln x$$
$$B_2(\boldsymbol{x}) = -x$$
$$C(\boldsymbol{x}) = 0$$
$$D(\boldsymbol{\theta}) = D(\alpha,\beta)$$
$$= \alpha \ln \beta - \ln \Gamma(\alpha)$$

Parametric form 2: $e^{\alpha \ln x - \beta x - \ln x + \alpha \ln \beta - \ln \Gamma(\alpha)}$

$A_1(\boldsymbol{\theta}) = A_1(\alpha, \beta) = \alpha$
$A_2(\boldsymbol{\theta}) = A_2(\alpha, \beta) = \beta$
$B_1(\boldsymbol{x}) = \ln x$
$B_2(\boldsymbol{x}) = -x$
$C(\boldsymbol{x}) = -\ln x$
$D(\boldsymbol{\theta}) = D(\alpha, \beta)$
$= \alpha \ln \beta - \ln \Gamma(\alpha)$

Canonical form: $A_j(\boldsymbol{\theta}) = \theta_j$

Poisson distribution:

$$f(x|\theta) = f(x|\mu) = \frac{\mu^x}{x!} e^{-\mu} = e^{(\ln \mu)\cdot x - \ln x! - \mu} \quad ; \; x = 0,1,\ldots$$

$$\left(= e^{(\ln \mu)\cdot x - \ln \Gamma(x+1) - \mu}\right)$$

$A_1(\boldsymbol{\theta}) = A(\mu) = \ln \mu$
$B_1(\boldsymbol{x}) = B(x) = x$
$C(\boldsymbol{x}) = -\ln x\,!$
$D(\boldsymbol{\theta}) = D(\mu) = -\mu$

Univariate normal distribution:

$$f(x|\theta) = f\left(x\middle|\mu, \sigma^2\right) = (2\pi\sigma^2)^{-0.5} e^{-\frac{(x-\mu)^2}{2\sigma^2}} \quad ; \; -\infty < x < \infty$$

$$= e^{-\left(1/\left(2\sigma^2\right)\right)\cdot x^2 + \left(\mu/\sigma^2\right)\cdot x - 0.5\ \ln(2\pi) - \mu^2/\left(2\sigma^2\right) - 0.5\cdot\ln\sigma^2}$$

$$A_1(\boldsymbol{\theta}) = A_1(\mu, \sigma^2) = \frac{1}{2\sigma^2}$$

$$A_2(\boldsymbol{\theta}) = A_2(\mu, \sigma^2) = \frac{\mu}{\sigma^2}$$

$$B_1(\boldsymbol{x}) = -x^2$$

$$B_2(\boldsymbol{x}) = x$$

$$C(\boldsymbol{x}) = -0.5\ln(2\pi)$$

$$D(\boldsymbol{\theta}) = D(\mu, \sigma^2) = -\frac{\mu^2}{2\sigma^2} - 0.5 \cdot \ln\sigma^2$$

Bernoulli distribution:

$$f(x|\theta) = f(x|p) = p^x(1-p)^{1-x} \qquad ; \; x = 0,1$$

$$= e^{(\ln p)\cdot x - (\ln(1-p))\cdot x + \ln(1-p)} = e^{\left(\ln\left(\frac{p}{1-p}\right)\right)\cdot x + \ln(1-p)}$$

$$A_1(\boldsymbol{\theta}) = A(p) = \ln\left(\frac{p}{1-p}\right)$$
$$B_1(\boldsymbol{x}) = B(x) = x$$
$$C(\boldsymbol{x}) = 0$$
$$D(\boldsymbol{\theta}) = D(p) = \ln(1-p)$$

Exercise: The binomial distribution belongs to the exponential class if (conditioned on) the number of trials is fixed. Why?

Conjugate families of distributions when the likelihood belongs to the exponential class

pdf (or pmf) of sample point distribution : $f(\boldsymbol{x}|\boldsymbol{\theta}) = e^{\sum_{j=1}^{k} A_j(\boldsymbol{\theta})B_j(\boldsymbol{x})+C(\boldsymbol{x})+D(\boldsymbol{\theta})}$

Likelihood from sample of n observations:

$$\prod_{i=1}^{n} f(\boldsymbol{x}_i|\boldsymbol{\theta}) = \prod_{i=1}^{n} e^{\sum_{j=1}^{k} A_j(\boldsymbol{\theta})B_j(\boldsymbol{x}_i)+C(\boldsymbol{x}_i)+D(\boldsymbol{\theta})}$$

$$= e^{\sum_{i=1}^{n}\left(\sum_{j=1}^{k} A_j(\boldsymbol{\theta})B_j(\boldsymbol{x}_i)+C(\boldsymbol{x}_i)+D(\boldsymbol{\theta})\right)}$$

$$= e^{\sum_{j=1}^{k} A_j(\boldsymbol{\theta}) \underbrace{\sum_{i=1}^{n} B_j(\boldsymbol{x}_i)}_{B_j'(\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\})}+\underbrace{\sum_{i=1}^{n} C(\boldsymbol{x}_i)}_{C'(\{\boldsymbol{x}_1, \dots, \boldsymbol{x}_n\})}+n\cdot D(\boldsymbol{\theta})}$$

Hence the multivariate array $\{\boldsymbol{X}_1, \dots, \boldsymbol{X}_n\}$ with independent marginal distributions all with density $f(\boldsymbol{x} \mid \boldsymbol{\theta})$ also belongs to the exponential class.

Now, mimic the structure of the exponential class (for the marginal distributions or the likelihood) and define the prior density for $\boldsymbol{\theta}$ as

$$f'\left(\boldsymbol{\theta} \middle| \alpha_1, \dots, \alpha_k, \alpha_{k+1}\right)$$

$$= e^{\sum_{j=1}^{k} A_j(\boldsymbol{\theta})\cdot\alpha_j + \alpha_{k+1}\cdot D(\boldsymbol{\theta}) + K(\alpha_1,\dots,\alpha_k,\alpha_{k+1})}$$

$$\propto e^{\sum_{j=1}^{k} A_j(\boldsymbol{\theta})\cdot\alpha_j + \alpha_{k+1}\cdot D(\boldsymbol{\theta})}$$

where $\alpha_1, \dots, \alpha_{k+1}$ are the hyperparameters of this prior distribution and $K(\ )$ is a function of $\alpha_1, \dots, \alpha_{k+1}$ only .

Then the posterior becomes

$$f''\left(\boldsymbol{\theta}\middle|\{\boldsymbol{x}\},\alpha_1,\dots,\alpha_k,\alpha_{k+1}\right)=f''\left(\boldsymbol{\theta}\middle|\boldsymbol{x}_1,\dots,\boldsymbol{x}_n;\alpha_1,\dots,\alpha_k,\alpha_{k+1}\right)$$

$$\propto\underbrace{\prod_{i=1}^{n}f(\boldsymbol{x}_i|\boldsymbol{\theta})}_{\text{likelihood}}\cdot f'\left(\boldsymbol{\theta}\middle|\alpha_1,\dots,\alpha_k,\alpha_{k+1}\right)$$

$$=e^{\sum_{j=1}^{k}A_j(\boldsymbol{\theta})\sum_{i=1}^{n}B_j(\boldsymbol{x}_i)+\sum_{i=1}^{n}C(\boldsymbol{x}_i)+n\cdot D(\boldsymbol{\theta})}\cdot e^{\sum_{j=1}^{k}A_j(\boldsymbol{\theta})\cdot\alpha_j+\alpha_{k+1}\cdot D(\boldsymbol{\theta})+K(\alpha_1,\dots,\alpha_k,\alpha_{k+1})}$$

$$=e^{\sum_{i=1}^{n}C(\boldsymbol{x}_i)}e^{K(\alpha_1,\dots,\alpha_k,\alpha_{k+1})}e^{\sum_{j=1}^{k}A_j(\boldsymbol{\theta})\left(\sum_{i=1}^{n}B_j(\boldsymbol{x}_i)+\alpha_j\right)+(n+\alpha_{k+1})\cdot D(\boldsymbol{\theta})}$$

$$\propto e^{\sum_{j=1}^{k}A_j(\boldsymbol{\theta})\left(\sum_{i=1}^{n}B_j(\boldsymbol{x}_i)+\alpha_j\right)+(n+\alpha_{k+1})\cdot D(\boldsymbol{\theta})}$$

i.e. the posterior distribution is of the same form as the prior distribution but with hyperparameters

$$\alpha_1+\sum_{i=1}^{n}B_1(\boldsymbol{x}_i),\dots,\alpha_k+\sum_{i=1}^{n}B_k(\boldsymbol{x}_i),\alpha_{k+1}+n$$

instead of

$$\alpha_1,\dots,\alpha_k,\alpha_{k+1}$$

Some common families (within or outside the exponential family):

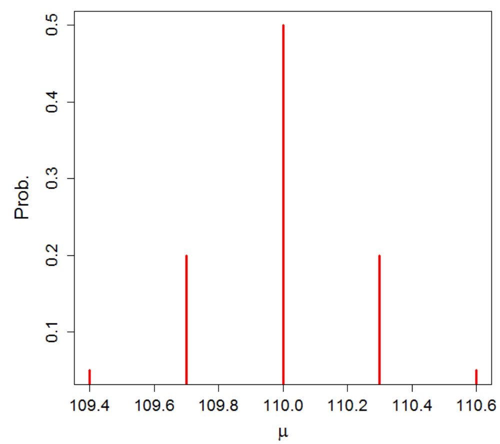
Conjugate prior	*Sample distribution*	*Posterior*
Beta	Binomial	Beta
$\pi \sim Beta(\alpha, \beta)$	$X \sim Bin(n, \pi)$	$\pi \vert x \sim Beta(\alpha + x, \beta + n - x)$
Normal	Normal, known σ^2	Normal
$\mu \sim N(\varphi, \tau^2)$	$X_i \sim N(\mu, \sigma^2)$	$\mu \vert \bar{x} \sim N\left(\frac{\sigma^2}{\sigma^2 + n\tau^2}\varphi + \frac{n\tau^2}{\sigma^2 + n\tau^2}\bar{x}, \frac{\sigma^2\tau^2}{\sigma^2 + n\tau^2}\right)$
Gamma	Poisson	Gamma
$\lambda \sim Gamma(\alpha, \beta)$	$X_i \sim Po(\lambda)$	$\lambda \vert \sum x_i \sim Gamma(\alpha + \sum x_i, \beta + n)$
Pareto	Uniform	Pareto
$p(\theta) \propto \theta^{-\alpha};\ \theta \geq \beta$	$X_i \sim U(0, \theta)$	$q(\theta; \mathrm{x}) \propto \theta^{-(\alpha+n)};\ \theta \geq \max(\beta, x_{(n)})$

Exercise 4.27

27. A production manager is interested in the mean weight of items turned out by a particular process. He feels that the weight of items from the process is normally distributed with mean $\tilde{\mu}$ and that $\tilde{\mu}$ is either 109.4, 109.7, 110.0, 110.3, or 110.6. The production manager assesses prior probabilities of $P(\tilde{\mu} = 109.4) = 0.05$, $P(\tilde{\mu} = 109.7) = 0.20$, $P(\tilde{\mu} = 110.0) = 0.50$, $P(\tilde{\mu} = 110.3) = 0.20$, and $P(\tilde{\mu} = 110.6) = 0.05$. From past experience, he is willing to assume that the process variance is $\sigma^2 = 4$. He randomly selects five items from the process and weighs them, with the following results: 108, 109, 107.4, 109.6, and 112. Find the production manager's posterior distribution and compute the means and the variances of the prior and posterior distributions.

Prior distribution of $\tilde{\mu}$:

$$\tilde{\mu} = \begin{cases} \mu_1 = 109.4 & \text{with prob. } 0.05\ (= p(\mu_1)) \\ \mu_2 = 109.7 & \text{with prob. } 0.20\ (= p(\mu_2)) \\ \mu_3 = 110.0 & \text{with prob. } 0.50\ (= p(\mu_3)) \\ \mu_4 = 110.3 & \text{with prob. } 0.20\ (= p(\mu_4)) \\ \mu_5 = 110.6 & \text{with prob. } 0.05\ (= p(\mu_5)) \end{cases}$$



Discretized normal distribution?

Data: $\boldsymbol{y}$ = {108.0 109.0 107.4 109.6 112.0} $\sim N(\tilde{\mu}, \sigma^2 \approx 4)$

Sample point density: $f(y|\tilde{\mu} = \mu) = (2\pi\sigma^2)^{-0.5} e^{-\frac{(y-\mu)^2}{2\sigma^2}}$; $\sigma^2 = 4$

Likelihood: $L(\mu; \boldsymbol{y}) = \prod_{j=1}^{n=5} f(y_j|\tilde{\mu} = \mu) = \prod_{j=1}^{n=5} (2\pi\sigma^2)^{-0.5} e^{-\frac{(y_j-\mu)^2}{2\sigma^2}}$

$$= (2\pi\sigma^2)^{-0.5n} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}(y_j-\mu)^2} = (8\pi)^{-2.5} e^{-\frac{1}{8}\sum_{j=1}^{5}(y_j-\mu)^2}$$

$$= (8\pi)^{-2.5} e^{-\frac{1}{8}[(108-\mu)^2+(109-\mu)^2+(107.4-\mu)^2+(109.6-\mu)^2+(112-\mu)^2]}$$

Posterior distribution of $\tilde{\mu}$:

$$f''(\mu|\boldsymbol{y}) = \frac{L(\mu;\boldsymbol{y})\cdot f'(\mu)}{\sum_{i=1}^{5} L(\mu_i;\boldsymbol{y})\cdot f'(\mu_i)} = \frac{(2\pi\sigma^2)^{-0.5n} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(y_j-\mu\right)^2} \cdot f'(\mu)}{\sum_{i=1}^{5}(2\pi\sigma^2)^{-0.5n} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(y_j-\mu_i\right)^2} \cdot f'(\mu_i)}$$

$$= \frac{e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(y_j-\mu\right)^2} \cdot f'(\mu)}{\sum_{i=1}^{5} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(y_j-\mu_i\right)^2} \cdot f'(\mu_i)} =$$

$$= \frac{e^{-\frac{1}{8}\sum_{j=1}^{n}\left(y_j-\mu\right)^2} \cdot f'(\mu)}{\begin{array}{l} e^{-\frac{1}{8}\cdot[(108-109.4)^2+(109-109.4)^2+(107.4-109.4)^2+(109.6-109.4)^2+(112-109.4)^2]} \cdot 0.05 + \\ e^{-\frac{1}{8}\cdot[(108-109.4)^2+(109-109.7)^2+(107.4-109.7)^2+(109.6-109.7)^2+(112-109.7)^2]} \cdot 0.20 + \\ e^{-\frac{1}{8}\cdot[(108-110.0)^2+(109-110.0)^2+(107.4-110.0)^2+(109.6-110.0)^2+(112-110.0)^2]} \cdot 0.50 + \\ e^{-\frac{1}{8}\cdot[(108-110.3)^2+(109-110.3)^2+(107.4-110.3)^2+(109.6-110.3)^2+(112-110.3)^2]} \cdot 0.20 + \\ e^{-\frac{1}{8}\cdot[(108-110.6)^2+(109-110.6)^2+(107.4-110.6)^2+(109.6-110.6)^2+(112-110.6)^2]} \cdot 0.05 \end{array}}$$

$$= \frac{e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(y_j-\mu\right)^2} \cdot f'(\mu)}{e^{-\frac{1}{8}\cdot 12.92}\cdot 0.05 + e^{-\frac{1}{8}\cdot 13.97}\cdot 0.20 + e^{-\frac{1}{8}\cdot 15.92}\cdot 0.50 + e^{-\frac{1}{8}\cdot 18.77}\cdot 0.20 + e^{-\frac{1}{8}\cdot 22.52}\cdot 0.05}$$

$$\approx \frac{e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(y_j-\mu\right)^2} \cdot f'(\mu)}{0.1353}$$

$\Rightarrow$

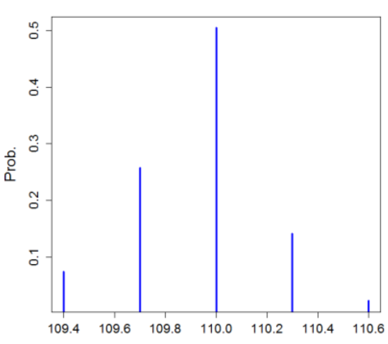
$$f''(\mu_1|\boldsymbol{y}) = \underline{f''(109.4|\boldsymbol{y})} \approx \frac{e^{-\frac{1}{8}[(108-109.4)^2+(109-109.4)^2+(107.4-109.4)^2+(109.6-109.4)^2+(112-109.4)^2]}\cdot 0.05}{0.1353} \approx \underline{0.0735}$$

$$f''(\mu_2|\boldsymbol{y}) = \underline{f''(109.7|\boldsymbol{y})} \approx \frac{e^{-\frac{1}{8}[(108-109.7)^2+(109-109.7)^2+(107.4-109.7)^2+(109.6-109.7)^2+(112-109.7)^2]}\cdot 0.20}{0.1353} \approx \underline{0.2578}$$

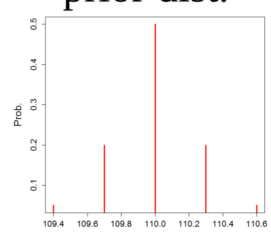
$$f''(\mu_3|\boldsymbol{y}) = \underline{f''(110.0|\boldsymbol{y})} \approx \frac{e^{-\frac{1}{8}[(108-110.0)^2+(109-110.0)^2+(107.4-110.0)^2+(109.6-110.0)^2+(112-110.0)^2]}\cdot 0.50}{0.1353} \approx \underline{0.5051}$$

$$f''(\mu_4|\boldsymbol{y}) = \underline{f''(110.3|\boldsymbol{y})} \approx \frac{e^{-\frac{1}{8}[(108-110.3)^2+(109-110.3)^2+(107.4-110.3)^2+(109.6-110.3)^2+(112-110.3)^2]}\cdot 0.20}{0.1353} \approx \underline{0.1415}$$

$$f''(\mu_5|\boldsymbol{y}) = \underline{f''(110.6|\boldsymbol{y})} \approx \frac{e^{\frac{1}{8}[(108-110.6)^2+(109-110.6)^2+(107.4-110.6)^2+(109.6-110.6)^2+(112-110.6)^2]}\cdot 0.05}{0.1353} \approx \underline{0.0221}$$



prior dist.



$$\underline{E_{\text{prior}}(\tilde{\mu})} = E(\tilde{\mu}) = 109.4 \cdot 0.05 + 109.7 \cdot 0.20 + 110.0 \cdot 0.50$$
$$+110.3 \cdot 0.20 + 110.6 \cdot 0.05 = \underline{110}$$

(obvious?)

$$\underline{Var_{\text{prior}}(\tilde{\mu})} = Var(\tilde{\mu}) = E(\tilde{\mu}^2) - \left(E(\tilde{\mu})\right)^2$$
$$= 109.4^2 \cdot 0.05 + 109.7^2 \cdot 0.20 + 110.0^2 \cdot 0.50$$
$$+110.3^2 \cdot 0.20 + 110.6^2 \cdot 0.05 - 110^2 = \underline{0.072}$$

$$\underline{E_{\text{posterior}}(\tilde{\mu})} = E(\tilde{\mu}|\boldsymbol{y})$$
$$= 109.4 \cdot 0.07348\ldots + 109.7 \cdot 0.25780\ldots + 110.0 \cdot 0.50508\ldots$$
$$+110.3 \cdot 0.14148\ldots + 110.6 \cdot 0.02213\ldots \approx \underline{109.9}$$

$$\underline{Var_{\text{posterior}}(\tilde{\mu})} = Var(\tilde{\mu}|\boldsymbol{y}) = E(\tilde{\mu}^2|\boldsymbol{y}) - \left(E(\tilde{\mu}|\boldsymbol{y})\right)^2$$
$$\approx 109.4^2 \cdot 0.07348\ldots + 109.7^2 \cdot 0.25780\ldots + 110.0^2 \cdot 0.50508\ldots$$
$$+110.3^2 \cdot 0.14148\ldots + 110.6^2 \cdot 0.02213\ldots - 109.7^2 \approx \underline{0.066}$$

Exercise 4.28

In Exercise 27, if $\tilde{\mu}$ is assumed to be continuous and if the prior distribution for $\tilde{\mu}$ is a normal distribution with mean 110 and variance 0.4, find the posterior distribution.

Prior distribution: $\tilde{\mu} \sim N(110,0.4) = N(m', \sigma'^2)$

Prior density: $p(\mu) = f'(\mu) = (2\pi\sigma'^2)^{-0.5} e^{-\frac{(\mu - m\prime)^2}{2\sigma\prime^2}} = (2\pi \cdot 0.4)^{-0.5} e^{-\frac{(\mu-110)^2}{0.8}}$

Data: $\mathbf{y}$ = {108.0 109.0 107.4 109.6 112.0} $\sim N(\tilde{\mu}, \sigma^2 \approx 4)$

$$\bar{y} = \frac{108 + 109 + 107.4 + 109.6 + 112}{5} = 109.2$$

$$s^2 = \frac{1}{4} \sum_{1}^{5} \left(y_j - 109.2\right)^2 = 3.18$$

$$f''(\mu|\boldsymbol{y}) = \frac{L(\mu;\boldsymbol{y})\cdot f'(\mu)}{\int_{-\infty}^{\infty} L(\mu;\boldsymbol{y})\cdot f'(\mu)d\mu} = \left\langle \begin{matrix} L(\mu;\boldsymbol{y}) \text{ from} \\ \text{Exercise } 4.27 \end{matrix} \right\rangle =$$

$$= \frac{(2\pi\sigma^2)^{-0.5n} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(y_j-\mu\right)^2} \cdot (2\pi\sigma'^2)^{-0.5} e^{-\frac{(\mu-m')^2}{2\sigma'^2}}}{\int_{-\infty}^{\infty}(2\pi\sigma^2)^{-0.5n} e^{-\frac{1}{2\sigma^2}\sum_{j=1}^{n}\left(y_j-\mu\right)^2} \cdot (2\pi\sigma'^2)^{-0.5} e^{-\frac{(\mu-m')^2}{2\sigma'^2}} d\mu} = \left\langle \begin{matrix} \text{"Completion} \\ \text{of squares"} \end{matrix} \right\rangle =$$

$$= (2\pi\sigma''^2)^{-0.5} e^{-\frac{(\mu-m'')^2}{2\sigma''^2}}$$

where

$$m'' = \frac{(1/\sigma'^2)\cdot m' + (n/\sigma^2)\cdot m}{(1/\sigma'^2) + (n/\sigma^2)} = \frac{(1/\sigma'^2)\cdot m' + (n/\sigma^2)\cdot \bar{y}}{(1/\sigma'^2) + (n/\sigma^2)}$$

$$= \frac{(1/0.4)\cdot 110 + (5/4)\cdot 109.2}{(1/0.4) + (5/4)} \approx 109.7$$

$$\sigma''^2 = \frac{\sigma^2\cdot\sigma'^2}{\sigma^2 + n\cdot\sigma'^2} = \frac{4\cdot 0.4}{4 + 5\cdot 0.4} \approx 0.267$$

Thus, the posterior distribution is $N(m''=109.7, \sigma''^2=0.267)$

Interpretation of prior distributions

Prior distribution for a proportion was taken up at meeting 2!

Prior distribution for the mean *of a population with continuous variation*

Very often we have reasons to work with normally distributed data to make inference about the population mean $\tilde{\mu}$.

If the population variance is (assumed to be) known = σ^2, we can – as was demonstrated in Exercise 4.28 – use the normal distribution as a conjugate prior distribution.

From sampling theory we know that – setting aside finite population corrections – the variance of the sample mean is the population variance divided by the sample size

$$Var\left(\bar{\tilde{y}} \middle| \sigma^2, n\right) = \frac{\sigma^2}{n}$$

If σ'^2 represents the prior variance of the unknown $\tilde{\mu}$ define a new parameter n' as

$$n' = \frac{\sigma^2}{\sigma'^2}$$

Hence, $$\sigma'^2 = \frac{\sigma^2}{n'}$$

This can be interpreted as the variance σ'^2 of a sample mean based on n' observations taken from the population with population variance σ^2.

n' then plays the role of the size of a virtual sample taken from the population on which the prior knowledge stems.

Note that it is not necessary for n' to be integer-valued, even if it often suffices to approximate with an integer.

For the prior and posterior distribution we may thus write

$$\tilde{\mu} \sim N\left(m', \frac{\sigma^2}{n'}\right) \qquad \tilde{\mu} | \boldsymbol{y} \sim N\left(m'', \frac{\sigma^2}{n''}\right) \qquad \text{where } n'' = \frac{\sigma^2}{\sigma''^2} = n' + n$$

Exercise 4.28 with alternative prior parametrization

Prior distribution: $\tilde{\mu} \sim N(110, 0.4) = N(m', \sigma'^2) = N(m', \sigma^2/10)$ since $\sigma^2 = 4$

$$f''(\mu|\boldsymbol{y}) = (2\pi\sigma''^2)^{-0.5} e^{-\frac{(\mu - m'')^2}{2\sigma''^2}}$$ [from the previous solution]

where

$$m'' = \frac{(1/\sigma'^2)\cdot m' + (n/\sigma^2)\cdot m}{(1/\sigma'^2) + (n/\sigma^2)} = \frac{(n'/\sigma^2)\cdot m' + (n/\sigma^2)\cdot m}{(n'/\sigma^2) + (n/\sigma^2)} = \left\langle \begin{array}{l}\text{All instances} \\ \text{of } 1/\sigma^2 \text{ can} \\ \text{be removed}\end{array} \right\rangle =$$

$$= \frac{n' \cdot m' + n \cdot m}{n' + n} = \frac{10 \cdot 110 + 5 \cdot 109.2}{10 + 5} \approx 109.7$$

$$\sigma''^2 = \frac{\sigma^2 \cdot \sigma'^2}{\sigma^2 + n \cdot \sigma'^2} = \frac{\sigma^2 \cdot (\sigma^2/n')}{\sigma^2 + n \cdot (\sigma^2/n')} = \frac{4 \cdot 0.4}{4 + 5 \cdot 0.4} = \frac{1.6}{6} \approx 0.267 = \frac{\sigma^2}{n''}$$

$$\Rightarrow n'' = \frac{\sigma^2}{1.6/6} = \frac{24}{1.6} = 15$$

And… $n'' = n' + n = 10 + 5$

A quick look at (an)other theory for understanding beliefs – part I

Consider the following case (from forensic science):

An attempt of burglary is recorded on a CCTV camera and it stands clear that the perpetrator is using a crowbar when trying to break the door to the premises (target of the intended burglary). The face of the perpetrator cannot be seen.

The perpetrator suddenly runs away leaving the crowbar behind him. Some time later the Police arrives to the crime scene and seizes the crowbar. Inspecting it more in detail reveals that it has a blue colour (crowbars sold are either painted – often in red or blue – or unpainted).

In the investigation interest is taken in a certain Mr Johnson, who is a well-reputed burglar. A visit is paid at his home, but he is not there. His wife – who opened the door - is asked whether Mr Johnson is in possession of a crowbar and what it looks like. She says he has a crowbar, and it is not painted.

What do we have here?

We have a crowbar, which we know was used for the burglary attempt thanks to the CCTV take-up.

Our question is: *Is it Mr Johnson's crowbar?*

To structure things:

Let A denote the statement "The crowbar belongs to Mr Johnson"

Let B denote "The crowbar is painted in blue"

Then we have a witness' statement: C = "Mr Johnson's crowbar is unpainted"

How do B and C influence our belief in A ?

A ="The crowbar belongs to Mr Johnson"
B = "The crowbar is blue"
C = "Witness says: Mr Johnson's crowbar is unpainted"

In terms of probabilities (using the subjective interpretation):

Why was Mr Johnson interesting from the beginning?

$P(A|I)$ must have been sufficiently high (where I is the background information available – before hearing what the witness (Mrs Johnson) said)

Is B relevant for A, i.e. is $P(A|B,I) \neq P(A|I)$?

Are A and B *conditionally dependent* given C,

i.e. is $P(A,B|C,I) \neq P(A|C,I) \cdot P(B|C,I)$?

There is a "problematic" difference between

C = "Witness says: Mr Johnson's crowbar is unpainted"

and (what may be confused with)

C' = "Mr Johnson's crowbar is unpainted"

A ="The crowbar belongs to Mr Johnson"
B = "The crowbar is blue"
C = "Witness says: Mr Johnson's crowbar is unpainted"
C' = "Mr Johnson's crowbar is unpainted"

For…

$P(A, B|C', I) = 0$ The crowbar cannot belong to Mr Johnson (*A*) and be blue (*B*) <u>if</u> Mr Johnson's crowbar is unpainted (*C'*)

but…

$P(A, B|C, I)$ is more difficult. In what way would the relevance between *A* and *B* be affected by a witness statement?

…and relevance with whom?

A ="The crowbar belongs to Mr Johnson"
B = "The crowbar is blue"
C = "Witness says: Mr Johnson's crowbar is unpainted"
C' = "Mr Johnson's crowbar is unpainted"

Decompose $P(A,B|C,I)$ using C' and $\neg C'$:

$$P(A,B|C,I) = P(A,B|C',C,I) \cdot P(C'|C,I) + P(A,B|\neg C',C,I) \cdot P(\neg C'|C,I) =$$

$$= 0 \cdot P(C'|C,I) + P(A,B|\neg C',C,I) \cdot P(\neg C'|C,I)$$

If $\neg C'$ holds, i.e. if Mr Johnson's crowbar *is* painted, then C' is no longer relevant (on its own) for A and B and we may write

$$P(A,B|C,I) = P(A|\neg C',I) \cdot \underbrace{P(B|\neg C',I)}_{\approx P(B|I)} \cdot \underbrace{P(\neg C'|C,I)}_{\text{Relates to the probability that the witness is lying}}$$

Hence, since $P(A,B|C,I) \neq P(A|C,I) \cdot P(B|C,I)$ A and B are conditionally dependent given C