MT4531/MT5731: (Advanced) Bayesian Inference

Conjugate Bayesian Analysis

Nicolò Margaritella

School of Mathematics and Statistics, University of St Andrews

–

University of St Andrews

Outline

Outline

Conjugate distributions

- **Definition:** A family of probability distributions, $\mathcal{F}$, is conjugate to a family of sampling distributions, $\mathcal{P}$, if whenever the prior belongs to the family, $\mathcal{F}$, then for any sample size and any value of observations, the posterior also belongs to the family, $\mathcal{F}$.
- Earlier examples of prior to posterior derivations were examples of conjugate analysis. For instance, when we assumed a **Gamma prior** for the parameter of the **Exponential distribution**.
- We will now see the case where **the prior is a Beta distribution**, and **the likelihood is the Binomial distribution**.

Conjugate distributions

- **Definition:** A family of probability distributions, $\mathcal{F}$, is conjugate to a family of sampling distributions, $\mathcal{P}$, if whenever the prior belongs to the family, $\mathcal{F}$, then for any sample size and any value of observations, the posterior also belongs to the family, $\mathcal{F}$.
- Earlier examples of prior to posterior derivations were examples of conjugate analysis. For instance, when we assumed a **Gamma prior** for the parameter of the **Exponential distribution**.
- We will now see the case where **the prior is a Beta distribution**, and **the likelihood is the Binomial distribution**.

Conjugate distributions

- **Definition:** A family of probability distributions, $\mathcal{F}$, is conjugate to a family of sampling distributions, $\mathcal{P}$, if whenever the prior belongs to the family, $\mathcal{F}$, then for any sample size and any value of observations, the posterior also belongs to the family, $\mathcal{F}$.
- Earlier examples of prior to posterior derivations were examples of conjugate analysis. For instance, when we assumed a **Gamma prior** for the parameter of the **Exponential distribution**.
- We will now see the case where **the prior is a Beta distribution**, and **the likelihood is the Binomial distribution**.

Outline

Beta prior - Binomial likelihood

- Suppose that a treatment (radiation) has a probability p of success in treating cancer. Success is denoted with $X = 1$, failure with $X = 0$.
- We monitor n randomly selected patients with (0 or 1) responses $x_1, x_2, ..., x_n$.
- We observe s positive responses in total, i.e.

$$\sum_{i=1}^{n} x_i = s.$$

- Suppose that we are prepared to assume that $x_1, \ldots, x_n$ are independently and identically distributed (iid), given p, with,

$$P(X_i = 1|p) = p, \quad i = 1, \ldots, n.$$

Beta prior - Binomial likelihood

- Suppose that a treatment (radiation) has a probability p of success in treating cancer. Success is denoted with $X = 1$, failure with $X = 0$.
- We monitor n randomly selected patients with (0 or 1) responses $x_1, x_2, ..., x_n$.
- We observe s positive responses in total, i.e.

$$\sum_{i=1}^{n} x_i = s.$$

- Suppose that we are prepared to assume that $x_1, \ldots, x_n$ are independently and identically distributed (iid), given p, with,

$$P(X_i = 1|p) = p, \quad i = 1, \ldots, n.$$

Beta prior - Binomial likelihood

- Suppose that a treatment (radiation) has a probability p of success in treating cancer. Success is denoted with $X = 1$, failure with $X = 0$.
- We monitor n randomly selected patients with (0 or 1) responses $x_1, x_2, \ldots, x_n$.
- We observe s positive responses in total, i.e.

$$\sum_{i=1}^{n} x_i = s.$$

- Suppose that we are prepared to assume that $x_1, \ldots, x_n$ are independently and identically distributed (iid), given p, with,

$$P(X_i = 1|p) = p, \quad i = 1, \ldots, n.$$

Beta prior - Binomial likelihood

- Suppose that a treatment (radiation) has a probability p of success in treating cancer. Success is denoted with $X = 1$, failure with $X = 0$.
- We monitor n randomly selected patients with (0 or 1) responses $x_1, x_2, ..., x_n$.
- We observe s positive responses in total, i.e.

$$\sum_{i=1}^{n} x_i = s.$$

- Suppose that we are prepared to assume that $x_1, \ldots, x_n$ are independently and identically distributed (iid), given p, with,

$$P(X_i = 1|p) = p, \quad i = 1, \ldots, n.$$

Likelihood

- $$P(X_1 = x_1, \ldots, X_n = x_n | p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^s(1-p)^{n-s}.$$
- Alternatively, we can assume that the total number of successes S in n patients follows a Binomial distribution so that, $S|p \sim Bin(n, p)$.
- The likelihood will then be,
$$p(S = s|p) = \binom{n}{s} p^s(1-p)^{n-s}.$$
- For the same prior, the posterior distribution **will be the same** under the two likelihoods; see next slide for a proof.
- (But the marginal distribution of the observations $f(\boldsymbol{x})$ will be different. See relevant question in Tutorial 2.)

Likelihood

- $$P(X_1 = x_1, ..., X_n = x_n|p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^s(1-p)^{n-s}.$$
- Alternatively, we can assume that the total number of successes S in n patients follows a Binomial distribution so that, $S|p \sim Bin(n, p)$.
- The likelihood will then be,
$$p(S = s|p) = \binom{n}{s} p^s(1-p)^{n-s}.$$
- For the same prior, the posterior distribution **will be the same** under the two likelihoods; see next slide for a proof.
- (But the marginal distribution of the observations $f(\boldsymbol{x})$ will be different. See relevant question in Tutorial 2.)

Likelihood

- $$P(X_1 = x_1, ..., X_n = x_n|p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^s(1-p)^{n-s}.$$
- Alternatively, we can assume that the total number of successes S in n patients follows a Binomial distribution so that, $S|p \sim Bin(n, p)$.
- The likelihood will then be,
$$p(S = s|p) = \binom{n}{s} p^s(1-p)^{n-s}.$$
- For the same prior, the posterior distribution **will be the same** under the two likelihoods; see next slide for a proof.
- (But the marginal distribution of the observations $f(\boldsymbol{x})$ will be different. See relevant question in Tutorial 2.)

Likelihood

- $$P(X_1 = x_1, ..., X_n = x_n | p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^s(1-p)^{n-s}.$$
- Alternatively, we can assume that the total number of successes S in n patients follows a Binomial distribution so that, $S|p \sim Bin(n, p)$.
- The likelihood will then be,
$$p(S = s|p) = \binom{n}{s} p^s(1-p)^{n-s}.$$
- For the same prior, the posterior distribution **will be the same** under the two likelihoods; see next slide for a proof.
- (But the marginal distribution of the observations $f(\boldsymbol{x})$ will be different. See relevant question in Tutorial 2.)

Likelihood

- $$P(X_1 = x_1, ..., X_n = x_n|p) = \prod_{i=1}^{n} p^{x_i}(1-p)^{1-x_i} = p^s(1-p)^{n-s}.$$
- Alternatively, we can assume that the total number of successes S in n patients follows a Binomial distribution so that, $S|p \sim Bin(n, p)$.
- The likelihood will then be,
$$p(S = s|p) = \binom{n}{s} p^s(1-p)^{n-s}.$$
- For the same prior, the posterior distribution **will be the same** under the two likelihoods; see next slide for a proof.
- (But the marginal distribution of the observations $f(\boldsymbol{x})$ will be different. See relevant question in Tutorial 2.)

Proof posteriors will be the same

- A (general) proof that the posterior distribution will be the same under the two likelihoods in the previous slide:
- Consider two experiments one yielding data x and the other y, so that $f(y|\theta) = cf(x|\theta)$ where c does not depend on θ.
- Then the two experiments contain identical information about θ, and lead to identical posterior distributions. (c cancels out in the posterior distribution calculations.)

$$\pi(\theta|y) = \frac{f(y|\theta)p(\theta)}{\int f(y|\theta)p(\theta)d\theta} = \frac{cf(x|\theta)p(\theta)}{\int cf(x|\theta)p(\theta)d\theta}$$

$$= \frac{f(x|\theta)p(\theta)}{\int f(x|\theta)p(\theta)d\theta} = \pi(\theta|x).$$

Proof posteriors will be the same

- A (general) proof that the posterior distribution will be the same under the two likelihoods in the previous slide:
- Consider two experiments one yielding data x and the other y, so that $f(y|\theta) = cf(x|\theta)$ where c does not depend on θ.
- Then the two experiments contain identical information about θ, and lead to identical posterior distributions. (c cancels out in the posterior distribution calculations.)

$$\pi(\theta|y) = \frac{f(y|\theta)p(\theta)}{\int f(y|\theta)p(\theta)d\theta} = \frac{cf(x|\theta)p(\theta)}{\int cf(x|\theta)p(\theta)d\theta}$$

$$= \frac{f(x|\theta)p(\theta)}{\int f(x|\theta)p(\theta)d\theta} = \pi(\theta|x).$$

Proof posteriors will be the same

- A (general) proof that the posterior distribution will be the same under the two likelihoods in the previous slide:
- Consider two experiments one yielding data x and the other y, so that $f(y|\theta) = cf(x|\theta)$ where c does not depend on θ.
- Then the two experiments contain identical information about θ, and lead to identical posterior distributions. (c cancels out in the posterior distribution calculations.)

$$\pi(\theta|y) = \frac{f(y|\theta)p(\theta)}{\int f(y|\theta)p(\theta)d\theta} = \frac{cf(x|\theta)p(\theta)}{\int cf(x|\theta)p(\theta)d\theta}$$

$$= \frac{f(x|\theta)p(\theta)}{\int f(x|\theta)p(\theta)d\theta} = \pi(\theta|x).$$

Conjugate Prior distribution

- We place a $Beta(a, b)$ prior on p, so that,

$$p(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} \propto p^{a-1}(1-p)^{b-1},$$

with Beta function $B(a, b) = \int_0^1 z^{a-1}(1-z)^{b-1}dz = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

- Note that $E(p) = \frac{a}{a+b}$ and $Var(p) = \frac{ab}{(a+b)^2(a+b+1)}$.

Conjugate Prior distribution

- We place a $Beta(a, b)$ prior on p, so that,

$$p(p) = \frac{1}{B(a,b)} p^{a-1}(1-p)^{b-1} \propto p^{a-1}(1-p)^{b-1},$$

  with Beta function $B(a,b) = \int_0^1 z^{a-1}(1-z)^{b-1}dz = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$
- Note that $E(p) = \frac{a}{a+b}$ and $Var(p) = \frac{ab}{(a+b)^2(a+b+1)}$.

Posterior distribution

- $$\pi(p|s) \propto f(s|p)p(p)$$

Posterior distribution

- $$\begin{aligned} \pi(p|s) &\propto f(s|p)p(p) \\ &\propto p^s(1-p)^{n-s} \times p^{a-1}(1-p)^{b-1} \\ &= p^{s+a-1}(1-p)^{n-s+b-1} \end{aligned}$$

- By inspection, we have that,

$$p|s \sim Beta(s+a, n-s+b).$$

- The posterior for the probability of success p is also a Beta distribution.
- So, the Beta distribution is a conjugate prior to the Binomial distribution.

Posterior distribution

- $$\begin{aligned} \pi(p|s) &\propto f(s|p)p(p) \\ &\propto p^s(1-p)^{n-s} \times p^{a-1}(1-p)^{b-1} \\ &= p^{s+a-1}(1-p)^{n-s+b-1} \end{aligned}$$
- By inspection, we have that,

$$p|s \sim Beta(s+a, n-s+b).$$

- The posterior for the probability of success p is also a Beta distribution.
- So, the Beta distribution is a conjugate prior to the Binomial distribution.

Posterior distribution

- $$\begin{aligned} \pi(p|s) &\propto f(s|p)p(p) \\ &\propto p^s(1-p)^{n-s} \times p^{a-1}(1-p)^{b-1} \\ &= p^{s+a-1}(1-p)^{n-s+b-1} \end{aligned}$$
- By inspection, we have that,
$$p|s \sim Beta(s+a, n-s+b).$$
- The posterior for the probability of success p is also a Beta distribution.
- So, the Beta distribution is a conjugate prior to the Binomial distribution.

Posterior distribution

- $$\begin{aligned}\pi(p|s) &\propto f(s|p)p(p)\\ &\propto p^s(1-p)^{n-s}\times p^{a-1}(1-p)^{b-1}\\ &= p^{s+a-1}(1-p)^{n-s+b-1}\end{aligned}$$

- By inspection, we have that,

$$p|s \sim Beta(s+a, n-s+b).$$

- The posterior for the probability of success p is also a Beta distribution.
- So, the Beta distribution is a conjugate prior to the Binomial distribution.

Examples of posterior distributions

- Using the R code *simpleR_BetaPrior_BinomialLikelihood.R* uploaded on Moodle, you can see what different Beta prior distributions look like, and the relative posterior distributions after a Binomial experiment is conducted. (See demonstration in lecture.)

Outline

Posterior distribution properties (1)

- To obtain insight into how the posterior combines information from the data and the prior...

$$\mathbb{E}_\pi(p) = \frac{s+a}{a+s+b+n-s} = \frac{s+a}{n+a+b}.$$

- We can rewrite this expectation in the form,

$$\mathbb{E}_\pi(p) = \frac{(a+b)\left(\frac{a}{a+b}\right) + n\left(\frac{s}{n}\right)}{n+a+b},$$

- which can be formulated as,

$$(1-w)\left(\frac{a}{a+b}\right) + w\left(\frac{s}{n}\right),$$

where $w = n/(n+a+b)$.

Posterior distribution properties (1)

- To obtain insight into how the posterior combines information from the data and the prior...

$$\mathbb{E}_\pi(p) = \frac{s+a}{a+s+b+n-s} = \frac{s+a}{n+a+b}.$$

- We can rewrite this expectation in the form,

$$\mathbb{E}_\pi(p) = \frac{(a+b)\left(\frac{a}{a+b}\right) + n\left(\frac{s}{n}\right)}{n+a+b},$$

- which can be reformulated as,

$$(1-w)\left(\frac{a}{a+b}\right) + w\left(\frac{s}{n}\right),$$

where $w = n/(n+a+b)$.

Posterior distribution properties (1)

- To obtain insight into how the posterior combines information from the data and the prior...

$$\mathbb{E}_\pi(p) = \frac{s+a}{a+s+b+n-s} = \frac{s+a}{n+a+b}.$$

- We can rewrite this expectation in the form,

$$\mathbb{E}_\pi(p) = \frac{(a+b)\left(\frac{a}{a+b}\right) + n\left(\frac{s}{n}\right)}{n+a+b},$$

- which can be reformulated as,

$$(1-w)\left(\frac{a}{a+b}\right) + w\left(\frac{s}{n}\right),$$

where $w = n/(n+a+b)$.

Posterior distribution properties (2)

- In other words, the Bayes estimate is a *weighted average* of the two quantities,

$$\frac{a}{a+b} \qquad \text{and} \qquad \frac{s}{n}.$$

- The first is the mean of the prior distribution and is the Bayes estimate we could use if we had no data.
- The latter is the classical estimate of p, derived via max. likelihood
- As the amount of data increases i.e. as n increases, more and more weight is placed on s/n;
- mathematically, in the limiting case, as $n \to \infty$, we have that $w \to 1$.
- Conversely, if we have no data, i.e. $n = 0$, then $w = 0$ and our only source of information on the parameter is contained within the prior.

Posterior distribution properties (2)

- In other words, the Bayes estimate is a *weighted average* of the two quantities,

$$\frac{a}{a+b} \qquad \text{and} \qquad \frac{s}{n}.$$

- The first is the mean of the prior distribution and is the Bayes estimate we could use if we had no data.
- The latter is the classical estimate of p, derived via max. likelihood
- As the amount of data increases i.e. as n increases, more and more weight is placed on s/n;
- mathematically, in the limiting case, as $n \to \infty$, we have that $w \to 1$.
- Conversely, if we have no data, i.e. $n = 0$, then $w = 0$ and our only source of information on the parameter is contained within the prior.

Posterior distribution properties (2)

- In other words, the Bayes estimate is a *weighted average* of the two quantities,

$$\frac{a}{a+b} \qquad \text{and} \qquad \frac{s}{n}.$$

- The first is the mean of the prior distribution and is the Bayes estimate we could use if we had no data.
- The latter is the classical estimate of p, derived via max. likelihood
- As the amount of data increases i.e. as n increases, more and more weight is placed on s/n;
- mathematically, in the limiting case, as $n \to \infty$, we have that $w \to 1$.
- Conversely, if we have no data, i.e. $n = 0$, then $w = 0$ and our only source of information on the parameter is contained within the prior.

Posterior distribution properties (2)

- In other words, the Bayes estimate is a *weighted average* of the two quantities,

$$\frac{a}{a+b} \qquad \text{and} \qquad \frac{s}{n}.$$

- The first is the mean of the prior distribution and is the Bayes estimate we could use if we had no data.
- The latter is the classical estimate of p, derived via max. likelihood
- As the amount of data increases i.e. as n increases, more and more weight is placed on s/n;
- mathematically, in the limiting case, as $n \to \infty$, we have that $w \to 1$.
- Conversely, if we have no data, i.e. $n = 0$, then $w = 0$ and our only source of information on the parameter is contained within the prior.

Posterior distribution properties (2)

- In other words, the Bayes estimate is a *weighted average* of the two quantities,

$$\frac{a}{a+b} \qquad \text{and} \qquad \frac{s}{n}.$$

- The first is the mean of the prior distribution and is the Bayes estimate we could use if we had no data.
- The latter is the classical estimate of p, derived via max. likelihood
- As the amount of data increases i.e. as n increases, more and more weight is placed on s/n;
- mathematically, in the limiting case, as $n \to \infty$, we have that $w \to 1$.
- Conversely, if we have no data, i.e. $n = 0$, then $w = 0$ and our only source of information on the parameter is contained within the prior.

Posterior distribution properties (2)

- In other words, the Bayes estimate is a *weighted average* of the two quantities,
$$\frac{a}{a+b} \qquad \text{and} \qquad \frac{s}{n}.$$
- The first is the mean of the prior distribution and is the Bayes estimate we could use if we had no data.
- The latter is the classical estimate of p, derived via max. likelihood
- As the amount of data increases i.e. as n increases, more and more weight is placed on s/n;
- mathematically, in the limiting case, as $n \to \infty$, we have that $w \to 1$.
- Conversely, if we have no data, i.e. $n = 0$, then $w = 0$ and our only source of information on the parameter is contained within the prior.

Posterior distribution properties (3)

- As the number of trials, n, increases, the precision of the posterior distribution for p increases, as we have more information.
- This can be seen formally, by considering the posterior variance for p,

$$Var_{\pi}(p) = \frac{(s+a)(n-s+b)}{(n+a+b)^2(n+a+b+1)}$$

- In the limiting case, as $n \to \infty$, we have that $Var_{\pi}(p) \to 0$.
- Thus, irrespective of our prior beliefs, as the amount of information increases, our posterior beliefs become more and more concentrated on a value of p tending to a value of s/n.

Posterior distribution properties (3)

- As the number of trials, n, increases, the precision of the posterior distribution for p increases, as we have more information.
- This can be seen formally, by considering the posterior variance for p,

$$Var_\pi(p) = \frac{(s+a)(n-s+b)}{(n+a+b)^2(n+a+b+1)}$$

- In the limiting case, as $n \to \infty$, we have that $Var_\pi(p) \to 0$.
- Thus, irrespective of our prior beliefs, as the amount of information increases, our posterior beliefs become more and more concentrated on a value of p tending to a value of s/n.

Posterior distribution properties (3)

- As the number of trials, n, increases, the precision of the posterior distribution for p increases, as we have more information.
- This can be seen formally, by considering the posterior variance for p,

$$Var_{\pi}(p) = \frac{(s+a)(n-s+b)}{(n+a+b)^2(n+a+b+1)}$$

- In the limiting case, as $n \to \infty$, we have that $Var_{\pi}(p) \to 0$.
- Thus, irrespective of our prior beliefs, as the amount of information increases, our posterior beliefs become more and more concentrated on a value of p tending to a value of s/n.

Posterior distribution properties (3)

- As the number of trials, n, increases, the precision of the posterior distribution for p increases, as we have more information.
- This can be seen formally, by considering the posterior variance for p,

$$Var_{\pi}(p) = \frac{(s+a)(n-s+b)}{(n+a+b)^2(n+a+b+1)}$$

- In the limiting case, as $n \to \infty$, we have that $Var_{\pi}(p) \to 0$.
- Thus, irrespective of our prior beliefs, as the amount of information increases, our posterior beliefs become more and more concentrated on a value of p tending to a value of s/n.

Outline

Normal prior - Normal likelihood (unknown mean, known variance)

- Assume that we observe conditionally independent observations $\boldsymbol{x} = \{x_1, \ldots, x_n\}$, drawn from the Normal distribution, i.e. given μ and σ, $X_i \overset{iid}{\sim} N(\mu, \sigma^2)$, $i = 1, \ldots, n$.
- Suppose that we specify the following prior on μ,

$$\mu \sim N(\phi, \tau^2).$$

Task: show that the posterior distribution for μ is,

$$\mu|\boldsymbol{x} \sim N\left(\frac{\tau^2 n\bar{x} + \sigma^2\phi}{\tau^2 n + \sigma^2}, \frac{\sigma^2\tau^2}{\tau^2 n + \sigma^2}\right).$$

Normal prior - Normal likelihood (unknown mean, known variance)

- Assume that we observe conditionally independent observations $\boldsymbol{x} = \{x_1, \ldots, x_n\}$, drawn from the Normal distribution, i.e. given μ and σ, $X_i \stackrel{iid}{\sim} N(\mu, \sigma^2)$, $i = 1, \ldots, n$.
- Suppose that we specify the following prior on μ,

$$\mu \sim N(\phi, \tau^2).$$

Task: show that the posterior distribution for μ is,

$$\mu|\boldsymbol{x} \sim N\left(\frac{\tau^2 n\bar{x} + \sigma^2\phi}{\tau^2 n + \sigma^2}, \frac{\sigma^2\tau^2}{\tau^2 n + \sigma^2}\right).$$

Normal prior - Normal likelihood (unknown mean, known variance)

- $$\mu|\boldsymbol{x} \sim N\left(\frac{\tau^2 n\bar{x} + \sigma^2\phi}{\tau^2 n + \sigma^2}, \frac{\sigma^2\tau^2}{\tau^2 n + \sigma^2}\right).$$
  Thus, the Normal prior on μ is a conjugate prior.
- It is also clear that the posterior mean is a mixture of the prior mean (ϕ) and the classical MLE for the mean ($\bar{x}$), as we can write,
  $$E(\mu|\boldsymbol{x}) = \frac{\tau^2 n\bar{x} + \sigma^2\phi}{\tau^2 n + \sigma^2} \quad = \quad w\bar{x} + (1-w)\phi,$$
  where,
  $$w = \frac{\tau^2 n}{\tau^2 n + \sigma^2}.$$

Normal prior - Normal likelihood (unknown mean, known variance)

- $$\mu|\boldsymbol{x} \sim N\left(\frac{\tau^2 n\bar{x} + \sigma^2\phi}{\tau^2 n + \sigma^2}, \frac{\sigma^2\tau^2}{\tau^2 n + \sigma^2}\right).$$

  Thus, the Normal prior on μ is a conjugate prior.
- It is also clear that the posterior mean is a mixture of the prior mean (ϕ) and the classical MLE for the mean ($\bar{x}$), as we can write,

$$E(\mu|\boldsymbol{x}) = \frac{\tau^2 n\bar{x} + \sigma^2\phi}{\tau^2 n + \sigma^2} \quad = \quad w\bar{x} + (1-w)\phi,$$

where,

$$w = \frac{\tau^2 n}{\tau^2 n + \sigma^2}.$$

Normal prior - Normal likelihood (unknown mean, known variance)

- $$\mu|\boldsymbol{x} \sim N\left(w\bar{x} + (1-w)\phi, \frac{\sigma^2\tau^2}{\tau^2 n + \sigma^2}\right); \quad w = \frac{\tau^2 n}{\tau^2 n + \sigma^2},$$

  The value of the prior variance, τ^2, specifies the informativeness of the prior.
- **(1)** τ^2 small: as $\tau^2 \to 0$, the mean of the distribution tends to ϕ and the posterior variance tends to 0. Thus, the prior dominates the posterior distribution.
- **(2)** τ^2 large: as $\tau^2 \to \infty$, the posterior mean for μ tends to $\bar{x}$. Additionally, the variance tends to σ^2/n.

Normal prior - Normal likelihood (unknown mean, known variance)

- $$\mu|\boldsymbol{x} \sim N\left(w\bar{x} + (1-w)\phi, \frac{\sigma^2\tau^2}{\tau^2 n + \sigma^2}\right); \quad w = \frac{\tau^2 n}{\tau^2 n + \sigma^2},$$

  The value of the prior variance, τ^2, specifies the informativeness of the prior.
- **(1)** τ^2 small: as $\tau^2 \to 0$, the mean of the distribution tends to ϕ and the posterior variance tends to 0. Thus, the prior dominates the posterior distribution.
- **(2)** τ^2 large: as $\tau^2 \to \infty$, the posterior mean for μ tends to $\bar{x}$. Additionally, the variance tends to σ^2/n.

Normal prior - Normal likelihood (unknown mean, known variance)

- $$\mu|\boldsymbol{x} \sim N\left(w\bar{x} + (1-w)\phi, \frac{\sigma^2\tau^2}{\tau^2 n + \sigma^2}\right); \quad w = \frac{\tau^2 n}{\tau^2 n + \sigma^2},$$

  The value of the prior variance, τ^2, specifies the informativeness of the prior.
- **(1)** τ^2 small: as $\tau^2 \to 0$, the mean of the distribution tends to ϕ and the posterior variance tends to 0. Thus, the prior dominates the posterior distribution.
- **(2)** τ^2 large: as $\tau^2 \to \infty$, the posterior mean for μ tends to $\bar{x}$. Additionally, the variance tends to σ^2/n.

Normal prior - Normal likelihood (unknown mean, known variance)

- **Task:** read Section 1.3 in the lecture notes and complete the relative exercise.