First easy examples with conjugate priors   Beta-Binomial-Modell

Analog: Billard-Kugel Versuch → Niete ≙ rechts der weißen Kugel

a) Lotterie: 10x gezogen, 9 Nieten und 1x Gewinn. Wie ist die Wahrscheinlichkeit, dass Verhältnis Nieten/Gewinn zwischen 8:1 und 10:1 liegt?

~~Frequentistisch~~ | Annahme unabhängige Ziehungen $X_i \sim B(\pi)$, $X = \Sigma_i X_i \sim B(10, \pi)$ n: Anzahl Ziehungen $f(x) = \binom{n}{x}\pi^x(1-\pi)^{n-x}$

Bereits Bayesianisch, weil Integr. über π

$$P\left(\frac{1}{11} < \pi < \frac{1}{9} \,\middle|\, x=1, n=10\right) = \frac{\int_{1/11}^{1/9}\binom{10}{1}\pi^1(1-\pi)^9 d\pi}{\int_0^1\binom{10}{1}\pi^1(1-\pi)^9 d\pi \,/\, 1/11!} \overset{R\,calc.}{=} \dots \frac{0.0078}{0.09091} \approx 0.086$$

Marginal distribution $f(k)$ discrete uniform

liegt an Beta-Prior, siehe unten

Was ist hier passiert? Implizit Integral über $\binom{n}{k}\pi^k(1-p)^{n-k}$ und unterstellen Gleich-Wahrscheinlichkeit für alle k $\Rightarrow \int_0^1\binom{n}{k}\pi^k(1-\pi)^{n-k}d\pi = \frac{1}{n+1}$ $\forall k = 0, \dots, n$

If we discretisize π in grid $\{0, 0.01, \dots, 0.99, 1\}$ then Prior $p(\pi) = \frac{1}{101}\mathbb{1}(\pi)$

1. Probability model $X \sim Bin(10, \pi) \to$ Likelihood $f(x|\pi) = \binom{n}{x}\pi^x(1-\pi)^{n-x}$

2. Prior distr. $\pi \sim U[0,1] = Beta(1,1)$ $p(\pi) = I_{[0,1]}(\pi)$

No knowledge, completely flat max. uninformative Prior ↑ Laplace-Prior

3. Posterior $p(\pi|x) = \frac{f(x|\pi)\,p(\pi)}{\int_\Pi f(x|\tilde\pi)p(\tilde\pi)d\tilde\pi} = \frac{\cancel{\binom{n}{x}}\pi^x(1-\pi)^{n-x}\cdot\frac{1}{B(a,b)}\pi^{a-1}(1-\pi)^{b-1}}{\cancel{\binom{n}{x}}\int_0^1\tilde\pi^x(1-\tilde\pi)^{n-x}\boxed{\frac{1}{B(a,b)}}\tilde\pi^{a-1}(1-\tilde\pi)^{b-1}d\tilde\pi}$

for general a, b, x   kann man vors Integral ziehen

$$\propto \underline{\underline{\pi^{a+x-1}(1-\pi)^{b+n-x-1}}} \text{ (posterior core density)}$$

$$= \frac{\cancel{\frac{1}{B(a,b)}}\pi^{a+x-1}(1-\pi)^{b+n-x-1}}{\cancel{\frac{1}{B(a,b)}}\int_0^1\pi^{a+x-1}(1-\pi)^{b+n-x-1}d\pi} = \frac{1}{B(a+x, b+n-x)}\pi^{a+x-1}(1-\pi)^{b+n-x-1}$$

$p(\pi; a, b) = \frac{1}{B(a,b)}\pi^{a-1}(1-\pi)^{b-1}$ with Beta Fct. $B(a,b) = \frac{\Gamma(a)\Gamma(b)}{\Gamma(a+b)}$

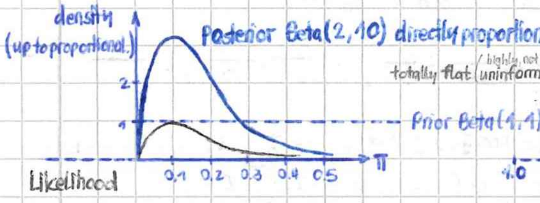
as $\int_0^1 t^{a-1}(1-t)^{b-1}dt$ "unabh. von π"

conjugate Prior Beta ⇒ Posterior also Beta-Distr.

Posterior distr. $\pi|x \sim Beta(a+x, b+n-x)$ (x: Σ success, $n-x$: Σ failure)

Here $a=b=1$, $n=10$, $x=1$ ⇓ cf Beta(2,10)

Use Distr. Fct for question above: $F(1/9) - F(1/11) \approx 0.086$



Posterior Beta(2,10) directly proportional to Likelihood due to flat Laplace-Prior

totally flat (uninformative (highly, not fully)) Prior, thus Posterior takes shape of (scaled) Likelihood $p(\pi|x=1) = \frac{f(x=1|\pi)\cdot p(\pi)}{f(x)} = \frac{\binom{n}{1}\pi^1(1-\pi)^{n-1}\cdot \overset{\pi\in[0,1]}{1}}{f(x)}$ für $\pi \in [0,1]$

very unlikely πs are where Posterior < Prior

1.0

.... Posterior Beta$(a+x, b+n-x)$ a↑ has same effect as x↑

Prior Beta(a,b): a interpretable as "prior Nr. of successes", b as "prior Nr. of failures"

⇒ a, b both small < 1 yields U-shape "Beta-density" ∪ improper prior b.c. $\int \dots d\pi = \infty$

By choosing constant flat prior $U[0,1] = Beta(1,1)$ we see connection between Frequentism and Bayesian Approach

Here: $\hat\pi_{MLE} = \frac{1}{10} = \hat\pi_{MAP}$ Max. Likelihood = Max/Mode of Posterior

MLE: $\ell(\pi) = \log\left(\binom{n}{x}\right) x\log\pi\,(n-x)\log(1-\pi)$ $\frac{\partial\ell(\pi)}{\partial\pi} \overset{!}{=} 0 \Rightarrow \frac{x}{\pi} - \frac{n-x}{1-\pi} = 0 \Leftrightarrow \frac{x(1-\pi)}{\pi} = n-x \Leftrightarrow x - x\pi = (n-x)\pi \Leftrightarrow \hat\pi_{ML} = \frac{x}{n}$

also take log here

Maximum-a-Posteriori: $\frac{\log(p(\pi|x))}{\partial\pi} \overset{!}{=}$ (ohne Konstante $\frac{1}{B(\dots)}$) $(a+x-1)\log\pi + (b+n-x-1)\log(1-\pi)$ $\frac{\partial\,\text{"}}{\partial\pi} = \frac{a+x-1}{\pi} - \frac{b+n-x-1}{1-\pi} \overset{!}{=} 0 \dots\dots \Leftrightarrow \hat\pi_{MAP} = \frac{a+x-1}{a+b+n-2} \overset{a=1, b=1}{=} \frac{x}{n}$ (1/10)

Posterior-EV: $\hat\pi_{PE} = E(\pi|x) \overset{\pi|x\sim Beta(a+x, b+n-x)}{=} \frac{a+x}{(a+x)+(b+n-x)} = \frac{a+x}{a+b+n} \overset{a=b=1}{=} \frac{x+1}{n+2}$ (1/6)   Modus/MAP here < Posterior-EV

most probable value for π < expectation for π on average

---

b) Let $y \overset{\text{Data Model}}{\sim} Poi(\lambda)$ and $\lambda \overset{\text{prior belief}}{\sim} Ga(\alpha, \beta)$ i.e. $p(\lambda; \alpha, \beta) = \frac{\beta^\alpha}{\Gamma(\alpha)}\cdot\lambda^{\alpha-1}\exp(-\lambda\beta) \propto \lambda^{\alpha-1}\exp(-\lambda\beta)$ Ga( ) as conjugate Prior

↳ Likelihood: $\overset{\text{iid}}{\prod_{i=1}^n}\lambda^{y_i}\exp(-\lambda)\cdot\frac{1}{y_i!} = \frac{1}{\prod_i y_i!}\lambda^{\Sigma_i y_i}\exp(-n\lambda) \propto \lambda^{\Sigma_i y_i}\exp(-n\lambda)$

as well as factor $\beta^\alpha/\Gamma(\alpha)$ in our prior

For Posterior we get proportionally, ignoring Bayesian Denominator and $\frac{1}{\prod y_i!}$ as both independ. of Param λ: $p(\lambda|y) \propto \lambda^{\Sigma y_i + \alpha - 1}\exp(-\lambda(n+\beta))$ (and write $n\bar y$)

With the knowledge that $Ga(\alpha,\beta)$ is a conjugate Prior for Poisson Distr, we conclude with posterior core that $\lambda|y \sim Ga(\Sigma y_i + \alpha, n+\beta)$ ($Poi(\lambda)$; $n\bar y$)

c) Let X be Erlang-distributed with unknown Param λ and known n Erlang density $f(x) = \frac{1}{(n-1)!}\lambda^n x^{n-1}\exp(-\lambda x)$, $X \sim Erlang(n, \lambda)$

Generalisation of Exp. Distr. for example as a sum of independent waiting times. **Gamma Conjugate Prior** not surprising as Erlang is special case of Gamma

Conjugate Prior is $\lambda \sim Ga(\alpha, \beta)$, so $p(\lambda; \alpha, \beta) \propto \lambda^{\alpha-1}\exp(-\beta\lambda)$ ⇒ $p(\lambda|x) \propto \lambda^n \underbrace{x^{n-1}}_{\text{indep of }\lambda}\exp(-\lambda x)\cdot\lambda^{\alpha-1}\exp(-\lambda\beta) \propto \lambda^{n+\alpha-1}\exp(-\lambda(x+\beta))$

$\lambda|x \sim Ga(\alpha+n, x+\beta)$

(*) Interesting: If $\mu_0 = 0$ and $s \to \infty$ we get $\mu|x \overset{a}{\sim} N(\bar x, \frac{\sigma^2}{n})$ Exchanging $\bar x$ and μ and viewing μ as fixed param. makes it look like frequentist CLT   Note that for $s \to \infty$ prior gets improper

d) Assume $X \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with known σ^2 Prior: $\mu \sim N(\mu_0, s)$ ⇒ $p(\mu) = \frac{1}{\sqrt{2\pi s}}\exp\left(-\frac{1}{2s}(\mu-\mu_0)^2\right) \propto \exp\left(-\frac{1}{2s}(\mu-\mu_0)^2\right)$ Conjugate

Jeffreys Prior: $p(\mu) \propto$ const. ($\sqrt{1/\sigma^2}$) edge case of conjugate core with $s \to \infty$ on $\mathbb{R}$, identical to Laplace

Likelihood: $f(x|\mu) \overset{i.i.d.}{=} \prod_{i=1}^n f(x_i|\mu) = \prod_{i=1}^n \frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right) = \frac{1}{(\sqrt{2\pi}\sigma)^n}\exp\left(-\frac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\mu)^2\right) \propto \exp\left(-\frac{1}{2\sigma^2}\Sigma_i(x_i-\mu)^2\right)$

Posterior: $p(\mu|x) \propto f(x|\mu)\cdot p(\mu) \propto \exp\left(-\frac{1}{2\sigma^2}\Sigma_i(x_i-\mu)^2\right)\exp\left(-\frac{1}{2s}(\mu-\mu_0)^2\right) = \exp\left(\underbrace{-\frac{1}{2}\cdot\frac{1}{\sigma^2}\Sigma_i(\cancel{x_i^2} - 2x_i\mu + \mu^2)}_{\propto -\frac{1}{2}\frac{n}{\sigma^2}\mu^2 + \frac{1}{\sigma^2}\Sigma_i x_i \mu} \underbrace{-\frac{1}{2}\cdot\frac{1}{s}(\mu^2 - 2\mu_0\mu + \cancel{\mu_0^2})}_{\propto -\frac{1}{2s}\mu^2 + \frac{\mu_0}{s}\mu}\right)$ (indep. of μ; all inside exp( ))

$$\propto \exp\left(-\frac{1}{2}\left(\frac{n}{\sigma^2} + \frac{1}{s}\right)\mu^2 + \left(\Sigma_i\frac{x_i}{\sigma^2} + \frac{\mu_0}{s}\right)\mu\right) = \exp\left(-\frac{1}{2}\left[\left(\frac{n}{\sigma^2} + \frac{1}{s}\right)\mu^2 - 2\left(\Sigma_i\frac{x_i}{\sigma^2} + \frac{\mu_0}{s}\right)\mu\right]\right) \underset{\text{quadr. extension}}{=} \exp\left(-\frac{1}{2}\underbrace{\left(\frac{n}{\sigma^2} + \frac{1}{s}\right)}_{\tilde s = 1/\dots}\left(\mu - \frac{\Sigma_i\frac{x_i}{\sigma^2} + \frac{\mu_0}{s}}{n/\sigma^2 + 1/s}\right)^2\right)$$

$\tilde s = \left(\frac{n}{\sigma^2} + \frac{1}{s}\right)^{-1} = \dots s\frac{\sigma^2}{n}/s + \frac{\sigma^2}{n}$ precision $1/\tilde s$

Overall $\mu|x \sim N\left(\underbrace{\tilde s\left(\frac{n\bar x}{\sigma^2} + \frac{\mu_0}{s}\right)}_{\tilde m}, \left(\frac{n}{\sigma^2} + \frac{1}{s}\right)^{-1}\right)$ (*)   What happens for $n \to \infty$?

Reformulated $\tilde m = \frac{s}{s + \frac{\sigma^2}{n}}\bar x + \frac{\sigma^2/n}{s + \frac{\sigma^2}{n}}\mu_0$ ≙ weighted mean of $\bar x$ (data-avg, MLE for μ) and prior-EV for μ

core of $N(\tilde m, \tilde s)$ distribution $\left(\frac{\sigma^2 s/n}{\sigma^2 s/n}\right)\tilde m = \dots \frac{s\bar x + \frac{\sigma^2}{n}\mu_0}{s + \sigma^2/n}$

Assume

e) $X \overset{i.i.d.}{\sim} N(\mu, \sigma^2)$ with known μ Show that conjugate Prior is $\sigma^2 \sim \text{InvGa}(a,b)$

Let prior be $\sigma^2 \sim \text{InvGa}(a,b)$ density: $p(\sigma^2) = \frac{b^a}{\Gamma(a)} (\sigma^2)^{-(a+1)} \exp\left(-\frac{b}{\sigma^2}\right)$

Data Likelihood: $L(\sigma^2 | x_1, \dots, x_z) = f(x|\sigma^2) = \prod_{i=1}^{n} (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(x_i-\mu)^2\right) = (2\pi\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n} (x_i-\mu)^2\right)$

Posterior proportional to Likelihood · Prior: $p(\sigma^2|x) \propto (\sigma^2)^{-n/2} \exp\left(-\frac{1}{2\sigma^2} \sum_{i=1}^{n}(x_i-\mu)^2\right) \cdot (\sigma^2)^{-(a+1)} \underbrace{\exp\left(-\frac{b}{\sigma^2}\right)}_{\text{pull inside}}$

$= (\sigma^2)^{-(n/2+a+1)} \exp\left(-\frac{1}{\sigma^2}\left(\frac{1}{2} \sum_{i=1}^{n}(x_i-\mu)^2 + b\right)\right)$ This is again core of an Inverse Gamma Distribution with updated parameters

$a' = a + n/2$, $b' = b + \frac{1}{2}\sum_i (x_i-\mu)^2$ which is $\frac{(b')^{a'}}{\Gamma(a')}$

$\Rightarrow \sigma^2 | x \sim \text{InvGa}\left(a + \frac{n}{2}, b + \frac{1}{2} \sum_{i=1}^{n} (x_i-\mu)^2\right)$ Posterior distribution, there was no need to derive specific form with Normalisierungskonstante

<u>Jeffreys Prior:</u> For one observation it is $f(x|\sigma^2) = (2\pi\sigma^2)^{-1/2} \exp\left(-\frac{1}{2\sigma^2}(x-\mu)^2\right)$ $\log(f(x|\sigma^2)) = -\frac{1}{2}\log(2\pi) - \frac{1}{2}\log(\sigma^2) - \frac{1}{2\sigma^2}(x-\mu)^2$

↑ same for $f(x|\mu) =$ "

If μ was unknown and σ^2 known (see page 1): $\frac{\partial}{\partial\mu}\log(f(x|\mu)) = -\frac{2}{2\sigma^2}\cdot(-1)(x-\mu) = \frac{1}{\sigma^2}(x-\mu)$, $\frac{\partial^2}{\partial\mu^2}$ " $= -\frac{1}{\sigma^2}$

Fisher-Info: $I(\mu) = -\mathbb{E}\left(\frac{\partial^2}{\partial\mu^2}\log(f(x|\mu))\right) = \frac{1}{\sigma^2} \rightarrow p^*(\mu) \propto I(\mu)^{1/2} = \underline{\underline{\sqrt{1/\sigma^2}}} \propto$ const. as σ^2 fixed

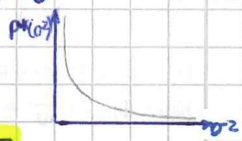
Now here for σ^2 unknown: $\frac{\partial}{\partial\sigma^2}\log(f(x|\sigma^2)) = -\frac{1}{2\sigma^2} + \frac{1}{2(\sigma^2)^2}(x-\mu)^2$, $\frac{\partial^2}{\partial(\sigma^2)^2}$ " $= \frac{1}{2(\sigma^2)^2} - \frac{1}{(\sigma^2)^3}(x-\mu)^2$

Knowing that $\mathbb{E}[(x-\mu)^2] = \sigma^2$ we get Fisher-Info: $I(\sigma^2) = -\mathbb{E}\left(\frac{\partial^2}{\partial(\sigma^2)^2}\log(f(x|\sigma^2))\right) = -\left(\frac{1}{2(\sigma^2)^2} - \frac{\sigma^2}{(\sigma^2)^3}\right) = -\left(\frac{1}{2(\sigma^2)^2} - \frac{2}{2(\sigma^2)^2}\right) = \frac{1}{2(\sigma^2)^2} = \frac{1}{2}\cdot\frac{1}{(\sigma^2)^2}$

↳ $I(\sigma^2) \propto \frac{1}{(\sigma^2)^2}$ and $p^*(\sigma^2) \propto I(\sigma^2)^{1/2} \propto \sqrt{\text{"}} = \underline{\underline{\frac{1}{\sigma^2}}}$ NOT const. like above

edge case of conjugate Prior $\sigma^2 \sim \text{InvGa}(a,b)$ with $a \to 0, b \to 0$

Unlike flat Jeffreys Prior for μ this Jeffreys Prior for σ^2 weighs small variances stronger than huge ones



Note] Its often easier for notation to use inverse variance / precision as parameter $\tau = (\sigma^2)^{-1} = \sigma^{-2} = \frac{1}{\sigma^2}$

This gives conjugate prior $\tau \sim \text{Ga}(a,b)$, Jeffreys Prior $p^*(\tau) \propto 1/\tau^2$

⇓

With that we get $p(\tau|x) \propto \tau^{n/2} \exp\left(-\frac{\tau}{2}\sum_{i=1}^{n}(x-\mu)^2\right) \cdot \tau^{a-1}\exp(-b\tau)$, so posterior $\tau | x \sim \text{Ga}\left(a + \frac{n}{2}, b + \frac{1}{2}\sum_i (x_i-\mu)^2\right)$

$= \tau^{a+n/2-1} \exp\left(-\tau\left(b + \frac{1}{2}\sum_i (x_i-\mu)^2\right)\right) \triangleq$ core of Ga(....., ...) density