Prior and Posterior Distributions:

Joint Distribution: $\psi(x,\theta) = \pi(\theta)\cdot f(x|\theta)$

denominator in the Baye's Theorem

Marginal Distribution: $m(x) = \int_\Theta \psi(x,\theta)\,d\theta = \int_\Theta \pi(\theta) f(x|\theta)\,d\theta$ [AKA: Prior predictive distribution]

Posterior Distribution: $\pi(\theta|x) = \dfrac{\psi(x,\theta)}{m(x)} = \dfrac{f(x|\theta)\,\pi(\theta)}{\int_\Theta \pi(\theta) f(x|\theta)\,d\theta} \propto f(x|\theta)\cdot\pi(\theta)$

Predictive Distribution:
$$\begin{cases} g(y|x) = \int_\Theta g(y|x,\theta)\cdot\pi(\theta|x)\,d\theta \overset{C.I.}{=\!=\!=} \int_\Theta g(y|\theta)\cdot\pi(\theta|x)\,d\theta & \text{[AKA: Posterior predictive function]} \\ g(y) = \int_\Theta g(y|\theta)\,\pi(\theta)\,d\theta & \text{[AKA: Prior predictive function]} \end{cases}$$

Beta-Binomial Model: [$x \sim Bin(n,\theta)$, $\theta \sim Beta(\alpha,\beta)$]

$f(x|\theta) = \binom{n}{x}\theta^x(1-\theta)^{n-x}$, $\pi(\theta) = \dfrac{1}{B(\alpha,\beta)}\theta^{\alpha-1}(1-\theta)^{\beta-1}$, $\psi(x,\theta) = \binom{n}{x}B(\alpha,\beta)^{-1}\theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}$

$m(x) = \int_\Theta \binom{n}{x}B(\alpha,\beta)^{-1}\theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}\,d\theta = \binom{n}{x}\dfrac{B(\alpha+x,\beta+n-x)}{B(\alpha,\beta)}$ [AKA: Beta-binomial distribution]

$\pi(\theta|x) \propto \psi(x,\theta) \propto \theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1} \sim Beta(\alpha+x,\beta+n-x)$

$f(y|x) = \int_\Theta \binom{m}{y}\theta^y(1-\theta)^{m-y}\cdot Beta(\alpha+x,\beta+n-x)^{-1}\,\theta^{\alpha+x-1}(1-\theta)^{\beta+n-x-1}\,d\theta$

$= \binom{m}{y}\cdot Be(\alpha+x,\beta+n-x)^{-1}\cdot Be(\alpha+x+y,\beta+n-x+m-y)$, $y \in \{0,1,2,\cdots,m\}$ [AKA: Beta-binomial]

Normal-Normal Model: [$x_i|\theta \overset{iid}{\sim} N(\theta,\sigma^2)$, $\theta \sim N(\mu,\tau^2)$, σ^2,μ,τ^2 are known]

$f(x_1,x_2,\cdots,x_n|\theta) = \prod_{i=1}^n \dfrac{1}{\sqrt{2\pi\sigma^2}}\cdot\exp\left(-\dfrac{(x_i-\theta)^2}{2\sigma^2}\right) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left(-\dfrac{1}{2\sigma^2}\sum_{i=1}^n(x_i-\bar{x}+\bar{x}-\theta)^2\right)$

$= (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left(-\dfrac{1}{2\sigma^2}\sum(x_i-\bar{x})^2 - \dfrac{1}{\sigma^2}\underbrace{\sum(x_i-\bar{x})(\bar{x}-\theta)}_{=0} - \dfrac{1}{2\sigma^2}\sum(\bar{x}-\theta)^2\right)$

$= (2\pi\sigma^2)^{-\frac{n}{2}}\exp\left(-\dfrac{1}{2\sigma^2}\sum(x_i-x)^2 - \dfrac{n}{2\sigma^2}(\bar{x}-\theta)^2\right)$

$\pi(\theta) = \dfrac{1}{\sqrt{2\pi\tau^2}}\exp\left(-\dfrac{(\theta-\mu)^2}{2\tau^2}\right)$, $\pi(\theta|x) \propto f(x|\theta)\cdot\pi(\theta) \propto \exp\left\{-\dfrac{n(\bar{x}-\theta)^2}{2\sigma^2} - \dfrac{(\theta-\mu)^2}{2\tau^2}\right\}$

$\propto \exp\left[-\frac{1}{2}\left\{(n/\sigma^2 + 1/\tau^2)\theta^2 - 2(n\bar{x}/\sigma^2 + \mu/\tau^2)\theta\right\}\right] = \exp\left[-\frac{1}{2}(\frac{n}{\sigma^2}+1/\tau^2)\cdot(\theta-\mu_1)^2\right]$

$\sim N(\mu_1,\tau_1^2) \Rightarrow \begin{cases}\mu_1 = (n\bar{x}/\sigma^2 + \mu/\tau^2)\cdot(n/\sigma^2+1/\tau^2)^{-1} \\ \tau_1^2 = (n/\sigma^2+1/\tau^2)^{-1}\end{cases}$

$f(y|x) = \int_\theta f(y|\theta)\cdot\pi(\theta|x)\,d\theta = \int_R \dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\dfrac{(y-\theta)^2}{2\sigma^2}\right)\cdot\dfrac{1}{\sqrt{2\pi\tau_1^2}}\exp\left(-\dfrac{(\theta-\mu_1)^2}{2\tau_1^2}\right)d\theta$

$= \dfrac{1}{\sqrt{4\pi^2\sigma^2\tau^2}}\int_R \exp\left[-\frac{1}{2}\left(\dfrac{y^2-2\theta y+\theta^2}{\sigma^2} + \dfrac{\theta^2-2\mu_1\theta+\mu_1^2}{\tau_1^2}\right)\right]d\theta = (2\pi\sigma\tau)^{-1}\int_R \exp\left[-\frac{1}{2}\left(\theta^2(\frac{1}{\sigma^2}+\frac{1}{\tau_1^2}) - 2\theta(\frac{y}{\sigma}+\frac{\mu_1}{\tau_1})\right)\right]d\theta\cdot\exp\left[-\frac{1}{2}\left(\frac{y^2}{\sigma^2}+\frac{\mu_1^2}{\tau_1^2}\right.\right.$

$= \dfrac{1}{\sqrt{2\pi(\sigma^2+\tau_1^2)}}\cdot\exp\left(-\frac{1}{2}(\sigma^2+\tau_1^2)^{-1}(y-\mu_1)^2\right)$ $\left.\begin{cases}\mu_1 = \dfrac{\frac{n\bar{x}}{\sigma^2}+\frac{\mu}{\tau^2}}{n/\sigma^2+1/\tau^2} \\ \tau_1^2 = \dfrac{1}{n/\sigma^2+1/\tau^2}\end{cases}\right] \Rightarrow y|x \sim N(\mu_1,\sigma^2+\tau_1^2)$

Exponential Family:

$f(x|\theta) = h(x)\cdot c(\theta)\cdot\exp(R(\theta)\cdot T(x))$ [$h(x)\geq 0$, $c(\theta)\geq 0$, $T(x)$ includes no θ, $R(\theta)$ includes no x]

Including: Normal, Gamma, Beta, binomial, Poisson, negative binomial.

Ex: Normal: $f(x|\mu,\sigma^2) = \dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\dfrac{1}{2\sigma^2}(x-\mu)^2\right) = \dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\dfrac{1}{2\sigma^2}(x^2-2\mu x+\mu^2)\right)$

$= \dfrac{1}{\sqrt{2\pi\sigma^2}}\exp\left(-\dfrac{\mu^2}{2\sigma^2}\right)\exp\left(-\dfrac{1}{2\sigma^2}x^2 + \dfrac{\mu}{\sigma^2}x\right)$

$h(x)=1$, $c(\theta) = \dfrac{1}{\sqrt{2\pi\sigma^2}}\cdot\exp\left(-\dfrac{\mu^2}{2\sigma^2}\right)$, $T(x) = \left(-\dfrac{x^2}{2}, x\right)$, $R(\theta) = \left(\dfrac{1}{\sigma^2}, \dfrac{\mu}{\sigma^2}\right)$

Improper Prior Distributions:

If $\int_\Theta \pi(\theta)d\theta = \infty$, it's called improper prior. Bayesian methods apply as long as the posterior is defined. And, posterior exists when pseudo marginal distribution $\int_\Theta f(x|\theta)\cdot\pi(\theta)\,d\theta$ is well defined ($<\infty$)

Sufficient Statistics:

Let $X_1, X_2, \cdots, X_n$ be a random sample of size n from a population and let $T(x_1, x_2, \cdots, x_n)$ be a real or vector valued function whose domain includes the sample space of $(x_1, x_2, \cdots, x_n)$. Then, the random variable or random vector $T(x_1, x_2, \cdots, x_n)$ is called a statistic. The probability distribution of $T(x_1, x_2, \cdots, x_n)$ is called sampling distribution. PS: $T(x)$ is said to be sufficient, if $x|T(x)$ does not depend on θ.

Fisher-Neyman Factorization Lemma: If $f(x|\theta) = g(T(x)|\theta)\cdot h(x|T(x))$

If $T(x) = T(y) \Rightarrow$ they must lead to the same inference $\ell(\theta) \propto g(T(x),\theta)$ [Sufficiency Principle]

PS: Inverse Gamma: $f(y|\alpha,\beta) = \frac{\beta^\alpha}{\Gamma(\alpha)} y^{-(\alpha+1)}\exp(-\beta\cdot\frac{1}{y})$

Example: $x_1, x_2, \cdots, x_n \overset{iid}{\sim} N(\theta, \sigma^2)$. μ and σ^2 are unknown, consider $\pi(\theta,\sigma) = 1/\sigma^2$ (Improper) [μ is θ here]

$f(x|\mu,\sigma^2) = \prod_{i=1}^n (2\pi\sigma^2)^{-\frac{1}{2}}\exp(-\frac{(x_i-\mu)^2}{2\sigma^2}) \propto (\sigma^2)^{-\frac{n}{2}}\exp(-\frac{S^2}{2\sigma^2} - \frac{n(\bar{x}-\theta)^2}{2\sigma^2})$ $\quad S^2 = \sum_{i=1}^n (x_i-\bar{x})^2,\ \bar{x} = \frac{\sum_{i=1}^n x_i}{n}$

$\pi(\mu,\sigma^2|x) = \pi(\mu,\sigma^2|\bar{x},s^2) \propto f(x|\mu,\sigma^2)\cdot\pi(\mu,\sigma^2) \propto (\sigma^2)^{-\frac{n}{2}-1}\exp(\frac{-s^2}{2\sigma^2} - \frac{n(\bar{x}-\mu)^2}{2\sigma^2})$ [$m(x)<\infty$, posterior exists]

$\pi(\mu|x,\sigma^2) \propto \pi(\mu,\sigma^2|x)$ since $\pi(\mu|x,\sigma^2) = \frac{\pi(\mu,\sigma^2|x)}{\int_\mu \pi(\mu,\sigma^2|x)d\mu} \propto \pi(\mu,\sigma^2|x)$ since $\int_\mu \pi(\mu,\sigma^2|x)d\mu = g(\sigma^2)$ is given

$\hookrightarrow \propto \exp(-\frac{n(\bar{x}-\mu)^2}{2\sigma^2}) \Rightarrow \mu|x,\sigma^2 \sim N(\bar{x}, \sigma^2/n)$

$\pi(\sigma^2|x) = \int_R \pi(\mu,\sigma^2|x)d\mu \propto (\sigma^2)^{-\frac{n}{2}-1}\exp(-\frac{s^2}{2\sigma^2})\int_R \exp(-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2)d\mu \propto (\sigma^2)^{-\frac{n}{2}-\frac{1}{2}}\exp(-\frac{s^2}{2\sigma^2})$

$\hookrightarrow$ It's a kernel of $IG(\frac{n}{2}-\frac{1}{2}, \frac{s^2}{2})$

$\pi(\mu|x) \overset{\times}{=} \underline{\int_0^\infty \pi(\mu,\sigma^2|x)d\sigma^2 \propto \exp(-\frac{n}{2\sigma^2}(\bar{x}-\mu)^2)\cdot\int_0^\infty \underbrace{(\sigma^2)^{-(\frac{n}{2}+1)}}_{\text{Kernel of } IG(\frac{n}{2},\frac{s^2}{2})}\cdot\exp(-\frac{s^2}{2\sigma^2})\,d\sigma^2}$ $\to$ Typical error! σ^2 included.

$= \int_0^\infty (\sigma^2)^{-(\frac{n}{2}+1)}\exp(-\frac{1}{2\sigma^2}[n(\bar{x}-\mu)^2+s^2])\,d\sigma^2 \Rightarrow$ Kernel of $IG(\frac{n}{2}, \frac{n(\bar{x}-\mu)^2+s^2}{2})$

$\propto (\frac{n(\bar{x}-\mu)}{2}+\frac{s^2}{2})^{-\frac{n}{2}} \propto (1+\frac{n(\bar{x}-\mu)^2}{s^2})^{-\frac{(n-1)+1}{2}}$ $\quad$ For t: $f(x) = \frac{\Gamma(\frac{V+1}{2})}{\Gamma(V/2)\sqrt{V\pi}\,\sigma}(1+\frac{1}{V}\cdot(\frac{x-\mu}{\sigma})^2)^{-\frac{V+1}{2}}$

$\Rightarrow \sigma^2 = \frac{s^2}{n(n-1)}\quad \mu = \bar{x},\ V = n-1.\quad \pi(\mu|x) \sim t(\bar{x}, \frac{s^2}{n(n-1)}, n-1)$

Example: $\pi_1(\theta|\sigma^2) \sim N(\mu, \sigma^2/n_0)$, $\pi_2(\sigma^2)$: $IG(V/2, S_0^2/2)$

Prior: $\pi(\theta,\sigma^2) = \frac{1}{\sqrt{2\pi\frac{\sigma^2}{n_0}}}\cdot\exp(-\frac{(\theta-\mu)^2}{2\frac{\sigma^2}{n_0}})\cdot\frac{(\frac{S_0^2}{2})^{V/2}}{\Gamma(V/2)}\cdot(\sigma^2)^{-\frac{V}{2}-1}\exp(-\frac{S_0^2}{2}\cdot\frac{1}{\sigma^2})$

$f(x|\theta,\sigma^2) = (2\pi\sigma^2)^{-\frac{n}{2}}\exp(-\frac{1}{2\sigma^2}[s^2+n(\bar{x}-\theta)^2])$: $S^2 = \sum_{i=1}^n (x_i-\bar{x})^2$, $\bar{x} = \sum_{i=1}^n x_i/n$.

$\pi(\theta,\sigma^2|x) \propto (\sigma^2)^{-\frac{n}{2}-\frac{V}{2}-1-\frac{1}{2}}\exp(-\frac{1}{2\sigma^2}[n_0(\theta-\mu)^2+S_0^2+s^2+n(\bar{x}-\theta)^2])$

$\pi(\theta|\sigma^2,x) \propto \pi(\theta,\sigma^2|x) \propto \exp(-\frac{1}{2\sigma^2}[(n_0+n)\theta^2 - 2\theta(n_0\mu+n\bar{x})] \sim N(\frac{n_0\mu+n\bar{x}}{n_0+n}, \frac{\sigma^2}{n_0+n})$

$\pi(\sigma^2|x) \propto \int_R \pi(\theta,\sigma^2|x)d\theta \propto (\sigma^2)^{-\frac{n}{2}-\frac{V}{2}-1-\frac{1}{2}}\exp(-\frac{S_0^2+s^2}{2\sigma^2})\int_R \exp(-\frac{1}{2\sigma^2}((n_0+n)\theta^2-2\theta(n_0\mu+n\bar{x})+n_0\mu^2+n\bar{x}^2))d\theta$

$= (\sigma^2)^{-\frac{n}{2}-\frac{V}{2}-1-\frac{1}{2}}\exp(-\frac{S_0^2+s^2}{2\sigma^2})\int_R \exp(-\frac{n_0+n}{2\sigma^2}(\theta-\frac{n_0\mu+n\bar{x}}{n_0+n})^2)d\theta\,\exp(\frac{1}{2\sigma^2}(\frac{(n_0\mu+n\bar{x})^2}{n_0+n} - \frac{(n_0+n)(n_0\mu^2+n\bar{x}^2)}{n_0+n}))$

$\propto (\sigma^2)^{-\frac{n+V}{2}-1}\exp(\frac{1}{2\sigma^2}\cdot\frac{n_0^2\mu^2+2nn_0\mu\bar{x}+n^2\bar{x}^2}{n_0+n} - \frac{n_0^2\mu^2+nn_0\mu^2+nn_0\bar{x}^2+n^2\bar{x}^2}{n_0+n})\exp(-\frac{S^2+S_0^2}{2\sigma^2})$

$= (\sigma^2)^{-\frac{n+V}{2}-1}\exp(\frac{-1}{2\sigma^2}(\frac{nn_0}{n_0+n})\cdot(\mu-\bar{x})^2)\exp(-\frac{S^2+S_0^2}{2\sigma^2})$

$= (\sigma^2)^{-\frac{n+V}{2}-1}\exp(-\frac{1}{\sigma^2}(\frac{\frac{nn_0}{n_0+n}(\mu-\bar{x})^2+S^2+S_0^2}{2})) \sim IG(\frac{n+V}{2}, \frac{\frac{nn_0}{n_0+n}(\mu-\bar{x})^2+S^2+S_0^2}{2})$

$\pi(\theta|x) = \int_0^\infty \pi(\theta,\sigma^2|x)\,d\sigma^2 \propto \int_0^\infty (\sigma^2)^{-\frac{n}{2}-\frac{v}{2}-1-\frac{1}{2}} \exp\left(-\frac{1}{2\sigma^2}[n_0(\theta-\mu)^2+S_0^2+S^2+n(\bar{x}-\theta)^2]\right)d\sigma^2$

It's a kernel of $IG\left(\frac{n+v+1}{2}, \frac{n_0(\theta-\mu)^2+n(\bar{x}-\theta)^2+S^2+S_0^2}{2}\right)$,

$$\propto \left[\frac{n_0(\theta-\mu)^2+n(\bar{x}-\theta)^2+S^2+S_0^2}{2}\right]^{-\frac{n+v+1}{2}}$$

$$\propto \left[\underbrace{S^2+S_0^2+\frac{nn_0}{n_0+n}(\mu-\bar{x})^2}_{S_1^2}+(n_0+n)\cdot\left(\theta-\frac{n_0\mu+n\bar{x}}{n_0+n}\right)^2\right]^{-\frac{n+v+1}{2}}$$

$$\propto \left(1+\frac{n_0+n}{S_1^2}\left(\theta-\left(\frac{n_0\mu+n\bar{x}}{n_0+n}\right)^2\right)\right)^{-\frac{n+v+1}{2}} = \left(1+\frac{\left(\theta-\frac{n_0\mu+n\bar{x}}{n_0+n}\right)^2}{S_1^2/(n_0+n)(n+v)}\cdot\frac{1}{n+v}\right)^{-\frac{n+v+1}{2}}$$

$\Rightarrow \mu_t = \frac{n_0\mu+n\bar{x}}{n_0+n}$, $\sigma_t^2 = \frac{S_1^2}{(n_0+n)(n+\mu)}$, $df = n+v$, it's a t-distribution.

Statistical Decision Theory:

Loss function: $\forall$ function L from $\Theta\times D$ in $[0,\infty)$: It evaluates the penalty $L(\theta,d)$ associated with the decision d, when the parameter takes value θ for all $(\theta,d)\in\Theta\times D$. Utility = − Loss.

Three necessary factors needing rigorous determination:

1. Distribution family for observations, $f(x|\theta)$ for $x\in\mathcal{X}$
2. Prior distribution for parameter $\pi(\theta)$, $\theta\in\Theta$
3. Loss association with the decisions, $L(\theta,\delta)\in[0,\infty)$

Conditionally on the observed data.

Posterior Expectation Loss:

$\rho(\pi,d|x) = E^\pi[L(\theta,d)|x] = \int_\Theta L(\theta|d)\cdot\pi(\theta|d)\,d\theta$: Overall Loss among all θ under decision δ.

↳ Bayes decision: $\delta^\pi(x)$ is any decision $d\in D$ which minimizes $\rho(\pi,d|x)$ *: Posterior!

Frequentist Risk (Average Risk)

$R(\theta,\delta) = E_\theta[L(\theta,\delta(x))] = \int_{\mathcal{X}} L(\theta,\delta(x))\cdot f(x|\theta)\,dx$: A function of θ; Overall loss among all x.

PS: $R(\theta,\delta) = E(L(\theta,\delta))$ if L is square loss function: $R(\theta,\delta) = E((\theta-\delta)^2) = E\{([\theta-E(\delta)]+[E(\delta)-\delta])^2\}$

$= E((\theta-E(\delta))^2) + E((\delta-E(\delta))^2) = \text{bias}^2 + Var(\delta)$.

$\rho(\pi,d|x) = E((\theta-d)^2) = d^2-2dE(\theta)+E(\theta^2)$, gets its minimum on $\hat{d}=E(\theta|x)$ → Posterior mean.

Integrated Risk (Bayes Risk):

$r(\pi,\delta) = E^\pi[R(\theta,\delta)] = \int_\Theta\int_{\mathcal{X}} L(\theta,d)f(x|\theta)\,dx\,\pi(\theta)\,d\theta$, it's a real number associated with estimator δ

$r(\pi,\delta) = \int_{\mathcal{X}}\rho(\theta,\delta)\,m(x)\,dx$: Because $f(x|\theta)\cdot\pi(\theta) = \pi(\theta|x)m(x)$

Bayes Estimator: It is associated with a prior and a loss function, and minimized $r(\pi,\delta)$.

Bayes Action : For every $x\in\mathcal{X}$, it's given by $\delta^\pi(x)$, $\arg\min_d \rho(\pi,d|x)$

How to choose decision:

1. Minimax Principle: $\sup_\theta R(\theta,\delta_1) < \sup_\theta(\theta,\delta_2)$, $\bar{R}$: Minimax Risk: $\inf_{\delta\in D}\sup E_\theta\{L(\theta,\delta(x))\}$
2. Admissibility: δ_0 is in admissible if $\exists\,\delta_1$ which dominates δ_0. i.e, for $\forall\theta$, $R(\theta,\delta_0)\geq R(\theta,\delta)$, and For at least on value of θ_0 of the parameter $R(\theta_0,\delta_0) > R(\theta_0,\delta)$. i.e. $R(\theta,\delta_0)\geqq R(\theta,\delta)$

PS: Bayes estimator satisfies admissibility automatically. i.e. If a prior distribution π is strictly positive on θ, with finite Bayes Risk and risk function $R(\theta,\delta)$ is a continuous function of θ, the δ^π is admissible.

PS: $L(\theta,d) = w(\theta)(\theta-d)^2 \Rightarrow \delta^\pi(x) = \frac{E^\pi(w(\theta)\cdot\theta|x)}{E^\pi(w(\theta)|x)}$

0-1 Loss: $\delta^\pi(x) = \begin{cases}1 & \text{if } P(\theta\in\Theta_0|x) > P(\theta\in\Theta_0|x)\\ 0 & \text{o.w.}\end{cases}$

Specification of Priors:

1. Subjective Determination: An ordering of relative likelihoods.
   When the parameter θ is finite, obtain a subjective evaluation of probabilities of different values of θ.
   When the Θ is noncountable, we may use histogram approach.
   Divide Θ into intervals, determine the subjective probability of each interval.
   Plot a probability histogram, if needed, smooth $\pi(\theta)$ can be sketched.
2. Parametric: Assume $\pi(\theta)$ is a functional form and choose the parameters (Hyper-) to control our belief.
3. Empirical Bayes: Assume the prior distribution of θ is in some parametric class with unknown parameters.
   And then use the data to specify the unknown parameters.
   Use MLE of data or M.M estimator to specify hyperparameters.