Distribution	pmf/pdf	Mean	Variance	Moment Generating Function: $\psi(t) = E[e^{tX}]$	Notes
$X \sim Bernoulli(p)$	$f_X(x) = p^x(1-p)^{1-x} * \mathbb{1}_{\{x=0,1\}}$	$E[X] = p$	$Var(X) = p(1-p)$	$\psi(t) = 1 - p + pe^t$	X =binary event w/ $P(success) = p$ $E[X] = E[X^2] = \cdots = E[X^k] = p$
$X \sim Binomial(n,p)$	$f_X(x) = \binom{n}{x} p^x(1-p)^{n-x} * \mathbb{1}_{\{x=0,1,\ldots,n\}}$	$E[X] = np$	$Var(X) = np(1-p)$	$\psi(t) = (1 - p + pe^t)^n$	X = # of successes in n trials.
$X \sim Poisson(\lambda)$ $\lambda > 0$	$f_X(x) = e^{-\lambda}\frac{\lambda^x}{x!} * \mathbb{1}_{\{x=0,1,\ldots\}}$	$E[X] = \lambda$	$Var(X) = \lambda$	$\psi(t) = \exp\{\lambda(e^t - 1)\}$	X = # of events occurring in a fixed time-period with a known rate λ .
$X \sim Geometric(p)$	$f_X(x) = (1-p)^x p * \mathbb{1}_{\{x=0,1,\ldots\}}$	$E[X] = \frac{1-p}{p}$	$Var(X) = \frac{1-p}{p^2}$	$\psi(t) = \frac{p}{1 - e^t(1-p)}$	X = # of unsuccessful trials preceding the first success
$X \sim Hypergeom(n,N,m)$	$f_X(x) = \frac{\binom{m}{x}\binom{N-m}{n-x}}{\binom{N}{n}}$	$E[X] = \frac{nm}{N}$	$Var(X) = \frac{nm}{N} * \frac{(N-m)(N-n)}{N(N-1)}$	–	N = population size , n = # of selections , m = # of successes . NO REPLACEMENT.
$X \sim NegativeBin(r,p)$	$f_X(x) = \binom{r+x-1}{x} p^r(1-p)^x * \mathbb{1}_{\{x=0,1,\ldots\}}$	$E[X] = \frac{r(1-p)}{p}$	$Var(X) = \frac{r(1-p)}{p^2}$	$\psi(t) = \left[\frac{p}{1-(1-p)e^t}\right]^r$	X = # of failures that occur before the r^{th} success.
$X \sim Uniform(a,b)$	$f_X(x) = \frac{1}{b-a} * \mathbb{1}_{\{a \le x \le b\}}$	$E[X] = \frac{a+b}{2}$	$Var(X) = \frac{(b-a)^2}{12}$	$\psi(t) = \frac{e^{tb} - e^{ta}}{t(b-a)}$	$Unif(0,1) \equiv Beta(1,1)$
$X \sim Exponential(\lambda)$ $\lambda > 0$	$f_X(x) = \lambda e^{-\lambda x} * \mathbb{1}_{\{x \ge 0\}}$	$E[X] = \frac{1}{\lambda}$	$Var(X) = \frac{1}{\lambda^2}$	$\psi(t) = \frac{\lambda}{\lambda - t} * \mathbb{1}_{\{t<\lambda\}}$	Memoryless Distribution: $P(X > s + t \mid X > s) = P(X > t)$
$X \sim Gamma(\alpha,\beta)$ $\alpha,\beta > 0$	$f_X(x) = \frac{\beta^\alpha e^{-\beta x} x^{\alpha-1}}{\Gamma(\alpha)} * \mathbb{1}_{\{x \ge 0\}}$	$E[X] = \frac{\alpha}{\beta}$	$Var(X) = \frac{\alpha}{\beta^2}$	$\psi(t) = \left(1 - \frac{t}{\beta}\right)^{-\alpha} * \mathbb{1}_{\{t<\beta\}}$	$X \sim Gamma(1,\beta)$ $\equiv Exponential(\beta)$ $\Gamma(\alpha) = (\alpha - 1)!$
$X \sim Beta(\alpha,\beta)$ $\alpha,\beta > 0$	$f_X(x) = \frac{1}{B(\alpha,\beta)} x^{\alpha-1}(1-x)^{\beta-1} * \mathbb{1}_{\{0 \le x \le 1\}}$	$E[X] = \frac{\alpha}{\alpha+\beta}$	$Var(X) = \frac{\alpha\beta}{(\alpha+\beta)^2(\alpha+\beta+1)}$	–	$B(\alpha,\beta) = \frac{\Gamma(\alpha)\Gamma(\beta)}{\Gamma(\alpha+\beta)}$
$X \sim Normal(\mu,\sigma^2)$ $\sigma^2 < 0$	$f_X(x) = \frac{1}{\sqrt{2\pi\sigma^2}} \exp\left\{-\frac{1}{2\sigma^2}(x-\mu)^2\right\} * \mathbb{1}_{\{x \in \mathbb{R}\}}$	$E[X] = \mu$	$Var(X) = \sigma^2$	$\psi(t) = \exp\left\{\mu t + \frac{\sigma^2 t^2}{2}\right\}$	–
$X \sim \chi^2_m$ CHI-SQUARE DIST. m degrees of freedom.	$f_X(x) = \frac{1}{2^{m/2}\Gamma(m/2)} x^{(m/2)-1} e^{-(x/2)} * \mathbb{1}_{\{x>0\}}$	$E[X] = m$	$Var(X) = 2m$	$\psi(t) = (1-2t)^{-m/2} * \mathbb{1}_{\{t<1/2\}}$	$Gamma\left(\alpha = \frac{m}{2}, \beta = \frac{1}{2}\right) \equiv \chi^2_m$ $Exponential\left(\beta = \frac{1}{2}\right) \equiv \chi^2_2$
$X \sim t_m$ t-DISTRIBUTION	$f_X(x) = \frac{\Gamma\left(\frac{m+1}{2}\right)}{(m\pi)^{1/2}\Gamma(m/2)} \left(1 + \frac{x^2}{m}\right)^{-(m+1)/2} * \mathbb{1}_{\{x \in \mathbb{R}\}}$	–	$Var(X) = \frac{m}{m-2}$, $m > 2$	–	$Y \sim \chi^2_m$ and $Z \sim N(0,1)$: $X = \frac{Z}{\sqrt{Y/m}} \sim t_m$

Prior : $\xi(\theta)$	Likelihood : $f_n(\mathbb{x} \mid \theta)$	Posterior : $\xi(\theta \mid \mathbb{X} = \mathbb{x})$
$\theta \sim Beta(\alpha,\beta)$	$X_1, \ldots, X_n \sim iid\ Bernoulli(\theta)$	$Beta\left(\alpha + \sum x_i, \beta + n - \sum x_i\right)$
$\theta \sim Gamma(\alpha,\beta)$	$X_1, \ldots, X_n \sim iid\ Poisson(\theta)$	$Gamma\left(\alpha + \sum x_i, \beta + n\right)$
$\theta \sim Gamma(\alpha,\beta)$	$X_1, \ldots, X_n \sim iid\ Exponential(\theta)$	$Gamma\left(\alpha + n, \beta + \sum x_i\right)$
$\theta \sim N(\mu_0, v_0^2)$	$X_1, \ldots, X_n \sim iid\ N(\theta,\sigma^2)$ $\sigma^2 > 0$ is known.	$N(\mu_1, v_1^2)$ $\mu_1 = \frac{\sigma^2\mu_0 + nv_0^2\bar{x}_n}{\sigma^2 + nv_0^2}$; $v_1^2 = \frac{\sigma^2 v_0^2}{\sigma^2 + nv_0^2}$

Theorem (7.2.1) : Suppose that the n random variables $X_1, \ldots, X_n$ form a random sample from a distribution for which the pdf is $f(x \mid \theta)$. Suppose also that the value of the parameter θ is unknown and the prior pdf of θ is $\xi(\theta)$. Then the posterior pdf is given by,

$$\xi(\theta \mid x_1, \ldots, x_n) = \frac{\xi(\theta) f(x_1, \ldots, x_n \mid \theta)}{g_n(x_1, \ldots, x_n)} \overset{iid}{=} \frac{\xi(\theta) \prod_{i=1}^n f(x_i \mid \theta)}{g_n(x_1, \ldots, x_n)} \quad \text{for } \theta \in \Omega \tag{1.1}$$

Where $g_n(x_1, \ldots, x_n)$ is the marginal joint pdf of $X_1, \ldots, X_n$. Furthermore, we can depict equation (1) equivalently:

$$\xi(\theta \mid x_1, \ldots, x_n) \propto f_n(x_1, \ldots, x_n \mid \theta)\xi(\theta) \tag{1.2}$$

Where the proportionality symbol $\propto$ is used to convey that the left-hand side is equal to the right-hand side except possibly up to a constant. The appropriate constant can be determined by using the fact that $\int_{\theta \in \Omega} \xi(\theta \mid x_1, \ldots, x_n) d\theta = 1$ or $g_n(x_1, \ldots, x_n) = \int_\Omega f_n(x_1, \ldots, x_n \mid \theta)\xi(\theta) d\theta$.

Definition (7.3.1) : ($Conjugate\ Family/Hyperparameters$) . Let $X_1, X_2, \ldots$ be conditionally iid given θ with common pdf $f(x \mid \theta)$. Let Ψ be a family of possible distributions over the parameter space Ω. Suppose that no matter which prior distribution ξ we choose from Ψ, no matter how many observations $\boldsymbol{X} = (X_1, \ldots, X_n)$ we observe, and no matter what are their observed values $\boldsymbol{x} = (x_1, \ldots, x_n)$, the posterior distribution $\xi(\theta \mid x_1, \ldots, x_n)$ is a member of Ψ. Then Ψ is called a *conjugate family of prior distributions under sampling* from the distributions $f(x \mid \theta)$. It is also said that the family Ψ is *closed under sampling* from the distributions $f(x \mid \theta)$. Finally, if the distributions in Ψ are parameterized by further parameters, then the associated parameters for the prior distribution are called the *prior hyperparameters* and the associated parameters of the posterior distribution are called the *posterior hyperparameters*.

Definition (7.3.2) : ($Improper\ Prior$) . Let ξ be a nonnegative function whose domain includes the parameter space of a statistical model. Suppose that $\int \xi(\theta) d\theta = \infty$. If we pretend as if $\xi(\theta)$ is the prior pdf of θ, then we are using an *improper prior* for θ.

Definition (7.4.3) : ($Bayes\ Estimator/Estimate$) . Let $L(\theta, a)$ be a loss function. For each possible value $x_1, \ldots, x_n$ of $X_1, \ldots, X_n$ let $\delta^*(x_1, \ldots, x_n)$ be a value of a such that $E[L(\theta, a) \mid x_1, \ldots, x_n] = \int_{\theta \in \Omega} L(\theta, a)\xi(\theta \mid x_1, \ldots, x_n) d\theta$ is minimized. Then δ^* is called a *Bayes estimator* of θ. Once $X_1 = x_1, \ldots, X_n = x_n$ is observed $\delta^*(x_1, \ldots, x_n)$ is called a *Bayes estimate* of θ.

Definition (7.4.4) : ($Squared\ Error\ Loss\ Function$) . The loss function $L(\theta, a) = (\theta - a)^2$ is called the *squared loss error*.

Corollary $(\mathbf{7.4.1})$: Let θ be a real-valued parameter. Suppose that the squared error loss function $L(\theta, a) = (\theta - a)^2$ is used and the posterior mean of θ, $E[\theta|X_1, \ldots, X_n]$, is finite. Then the Bayes estimator of θ is $\delta^*(X_1, \ldots, X_n) = E[\theta|X_1, \ldots, X_n]$.
Definition $(\mathbf{7.4.5})$: $(Absolute\ Error\ Loss\ Function)$. The loss function $L(\theta, a) = |\theta - a|$ is called the *absolute error loss*.
For every observed value $x_1, \ldots, x_n$, the Bayes estimate $\delta^*(x_1, \ldots, x_n)$ will now be the value of a for which the expectation $E[(|\theta - a|)|x_1, \ldots, x_n]$ is a minimum. This is when a is the median of the posterior distribution.

Consider a distribution for which the pdf is $f(x|\theta)$, where θ belongs to some parameter space Ω. It is said that the family of distributions obtained by letting θ vary over all values in Ω is an exponential family, if $f(x|\theta)$ can be written as follows for $\theta \in \Omega$ and all values of x:

$$f(x|\theta) = a(\theta)b(x)\, exp\{c(\theta)d(x)\}$$

Here $a(\theta)$ and $c(\theta)$ are arbitrary functions of θ, and $b(x)$ and $d(x)$ are arbitrary functions of x.