Reference Sheet - Willy Chen (Latest Version)

Common Distributions

Normal $X \sim N(\mu, \sigma^2)$

$$PDF: \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{1}{2}\frac{(x-\mu)^2}{\sigma^2})$$
$$MGF: exp(\mu t + \frac{\sigma^2 t^2}{2})$$

Lognormal $X \sim Lognormal(\mu, \sigma^2)$

$$PDF: \frac{1}{\sigma\sqrt{2\pi}} exp(-\frac{1}{2}\frac{(log(x)-\mu)^2}{\sigma^2}),\ x > 0$$
$$E[X] = exp(\mu + \frac{\sigma^2}{2}),\ Var(X) = [exp(\sigma^2) - 1]E[X]^2$$

Note: A lognormally distributed r.v. is an r.v. whose logged version is normally distributed.

Chi-Square $X \sim \chi^2_n$

Let $Z \sim Normal(0, I_n)$, $Z'Z = \sum_{i=1}^{n} Z_i^2 \sim \chi^2_n$

$$E[X] = n,\ Var(X) = 2n$$

t Distribution with $df = n$

Let $Z \sim Normal(0,1)$, $X \sim \chi^2_n$. Define $T \equiv \frac{Z}{\sqrt{X/n}}$.
Then $T \sim \mathcal{T}_n$. As $\lim_{n\to\infty} \mathcal{T}_n \to Normal(0,1)$

F Distribution with $df = n$

Let $X_1 \sim \chi^2_{k_1}$, $X_2 \sim \chi^2_{k_2}$. Define $W \equiv \frac{X_1/k_1}{X_2/k_2} \sim \mathcal{F}_{k_1,k_2}$

Gamma $X \sim Gamma(\alpha, \beta)$

$$PDF: \frac{1}{\beta^\alpha \Gamma(\alpha)} x^{\alpha-1} exp(-\frac{x}{\beta}),\ x > 0$$
$$MGF: (1-\beta t)^{-\alpha},\ t < \frac{1}{\beta}$$
$$E[X] = \alpha\beta,\ Var(X) = \alpha\beta^2$$

When $\alpha = 1$, this is equivalent to $Exponential(\frac{1}{\beta})$.
If $X, Y \sim Gamma(\alpha_0, \beta_0)$, $X + Y \sim Gamma(2\alpha_0, \beta_0)$
$Gamma(\alpha = \frac{n}{2}, \beta = 2) \equiv \chi^2_n$.
α represents the time waiting and β represents the scale of the event (e.g. $\frac{1}{\beta}$ customers come in every α hours, $\lambda = \frac{\beta}{\alpha}$ for exponential).
Note: This distribution is typically used to model a continuous time until an event. However, generally, the **gamma distribution is NOT memoryless** unless it is the case of an exponential distribution. In a general quesiton, try to use exponential instead (reducing α to 1. See problem $\star$ in selected problems for variations.

Exponential $X \sim Exponential(\lambda)$

$$PDF: \lambda e^{-\lambda x},\ \lambda > 0$$
$$CDF: 1 - e^{-\lambda x}$$
$$MGF: \frac{\lambda}{\lambda - t},\ t < \lambda$$
$$E[X] = \frac{1}{\lambda},\ Var(X) = \frac{1}{\lambda^2}$$

Note: This distribution is typically used to model a continuous time until an event. For an example, see problem $\star$ in selected problems.
Exponential is memoryless[1]

Binomial $X \sim Binomial(n, p)$

$$PMF: \binom{n}{k} p^k (1-p)^{n-k}$$
$$MGF: (1 - p + pe^t)^n$$
$$E[X] = np,\ Var(X) = np(1-p)$$

Negative Binomial $X \sim NegBin(\mu, \alpha)$

$$\Gamma(r) = \int_0^\infty exp(-u)u^{r-1}du,\ r > 0$$
$$\Gamma(k) = (k-1)!,\ k \in \mathbb{Z}_{++}$$
$$PMF: \frac{\Gamma(\alpha + x)}{\Gamma(\alpha)x!}\left(\frac{\alpha}{\alpha+\mu}\right)^\alpha \left(\frac{\mu}{\alpha+\mu}\right)^x,\ x \in \mathbb{Z}_+$$
$$MGF: \left(1 + \frac{\mu}{\alpha}[1 - exp(t)]\right)^{-\alpha},\ t < -ln\left(\frac{\mu}{\alpha+\mu}\right)$$
$$E[X] = \mu,\ Var(X) = \mu + \frac{\mu^2}{\alpha}$$

When $\alpha = 1$, this is the *geometric* distribution
As $\alpha \to \infty$, NB converges to $Poisson(\mu)$

Poisson $X \sim Poisson(\theta)$

$$PMF: \frac{exp(-\theta)\theta^x}{x!},\ x \in \mathbb{N} \cup \{0\}$$
$$CDF: exp(-\theta)\sum_{x=0}^{t} \frac{\theta^x}{x!}$$
$$MGF: exp[\theta(exp(t) - 1)]$$
$$E[X] = \theta,\ Var(X) = \theta$$

$Poisson(\theta_1) + Poisson(\theta_2) = Poisson(\theta_1 + \theta_2)$
Note: This distribution is typically used to model the probability of an event happening given a specific time period. λ is the frequency of the event in said time period.
Poisson is memoryless[1].

Geometric $X \sim Geometric(p)$

k total trials ($k \in \mathbb{N}$)

$$PMF: (1-p)^{k-1}p$$
$$CDF: 1 - (1-p)^{\lfloor k \rfloor}$$
$$MGF: \frac{pe^t}{1 - (1-p)e^t},\ t < -ln(1-p)$$
$$E[X] = \frac{1}{p},\ Var(X) = \frac{1-p}{p^2}$$

k failures before success ($k \in \mathbb{N} \cup \{0\}$)
This is the sepcial case of $\Gamma(1, \mu)$

$$PMF: (1-p)^k p$$
$$CDF: 1 - (1-p)^{\lfloor k \rfloor + 1}$$
$$MGF: \frac{p}{1 - (1-p)e^t},\ t < -ln(1-p)$$
$$E[X] = \frac{1-p}{p},\ Var(X) = \frac{1-p}{p^2}$$

Geometric is memoryless[1]

Some Common Use Cases

Continuous wait time before an event:
$\Gamma(\alpha, \beta)$ or $Exponential(\lambda) \equiv \Gamma(1, \frac{1}{\lambda}|$
Discrete wait time before an event: $NegBin(\mu, \alpha)$ or $Geometric(\lambda)$
Probability of event in a given time:
$Poisson(\theta)$

Important Properties

$\sigma-$algebra

Let Ω be the outcome space and $\mathcal{B}$ be the $\sigma-$algebra generated by $\mathcal{B}$. Then $\mathcal{B}$ must satisfy:

1. $\Omega \in \mathcal{B}$
2. $\forall A \in \mathcal{B},\ A^c \in \mathcal{B}$
3. $\forall i \in \mathbb{N}, A_i \in \mathcal{B},\ \bigcup_{i=1}^{\infty} A_i \in \mathcal{B}$

Probability of Random Draws

	Without Replacement	With Replacement
Ordered	$P^n_k = \frac{n!}{(n-k)!}$	n^k
Unordered	$C^n_k = \frac{n!}{(n-k)!k!}$	$C^{n+k-1}_k = \frac{(n+k-1)!}{k!(n-1)!}$

Bayes' Rule

$$P(A|B) = \frac{P(A \cap B)}{P(B)}$$
$$P(A|B)P(B) = P(B|A)P(A)$$
$$P(A|B) = \frac{P(B|A)P(A)}{P(B|A)P(A) + P(B|A^c)P(A^c)}$$

[1] For discrete $P(X > m + n \mid X \geq m) = P(X > n)$, for continuous $P(X > t + s \mid X > t) = P(X > s)$

Probability as Expectation

Define the indicator function $I\{statement\}$ to be

$$I\{statement\} \equiv \begin{cases} 1 & \text{Statement is TRUE} \\ 0 & \text{Statement is False} \end{cases}$$

Then the probability of an event is the expectation of the indicator function of the event happening:

$$P(A) = E[I\{A\}]$$

Markov's Inequality

$$P(h(X) \geq b) \leq \frac{E[h(X)]}{b}$$

Chebyshev's Inequality

For $c > 0$, $a > 0$, $E[X^2] < \infty$

$$P(|X-\mu| \geq c) \leq \frac{\sigma_X^2}{c^2}$$
$$P(|X-\mu| \geq a\sigma) \leq \frac{1}{a^2}$$

Cauchy-Schwartz Inequality

$$|E[XY]| \leq E[|XY|] \leq [E[X^2]]^{\frac{1}{2}}[E[Y^2]]^{\frac{1}{2}}$$

Jensen's Inequality

Let $\mathcal{X} = supp(X)$, if $g: \mathcal{X} \to \mathbb{R}$ is **convex**, then

$$g(E[X]) \leq E[g(X)]$$

Holder's Inequality

$\forall p, q \in [1, \infty)\, s.t.\, \frac{1}{p} + \frac{1}{q} = 1$

$$\|fg\|_1 = \|f\|_p \|g\|_q$$

Minkowski's Inequality

$\forall p \in [1, \infty)$,

$$E[|X+Y|^p]^{\frac{1}{p}} \leq E[|X|^p]^{\frac{1}{p}} + E[|Y|^p]^{\frac{1}{p}}$$
$$E[|X+Y|] \leq E[|X|] + E[|Y|]$$
$$SD(X+Y) \leq SD(X) + SD(Y)$$

Interesting Property of Expectation

$\forall X \geq 0,\ E[X] = \int\limits_{supp(X)} 1 - F(x)dx$

Law of Iterated Expectations

$$E_Y[Y] = E_X[E_Y[Y|X]] = E_X[E_Z[E_Y[Y|X,Z]|X]]$$

Law of Total Variance

$$Var(Y) = E[V[Y|X]] + V[E[Y|X]]$$

Conditional/Joint PDFs

$$\begin{aligned}
f_{XY}(x,y) = f_X(x) \cdot f_Y(y) &\iff X \perp\!\!\!\perp Y \\
f_X(x) &= \int_{supp(Y)} f_{XY}(x,y)dy \\
f_{Y|X} &= \frac{f_{XY}}{f_Y} = \frac{\int_{supp(Z)} f_{XYZ}dz}{f_Y} \\
&= \int_{supp(Z)} \frac{f_{XYZ}(x,y,z)}{f_Y(y)} \cdot \frac{f_{XY}(x,y)}{f_{XY}(x,y)}dz \\
&= \int_{supp(Z)} f_{Z|X,Y} \cdot f_{X|Y}dz
\end{aligned}$$

Moreover,

$$\begin{aligned}
f_{Y,X|Z} &= \frac{f_{YXZ}(y,x,z)}{f_Z(z)} = \frac{f_{Y|X,Z}(y|x,z)f_{X,Z}(x,z)}{f_Z(z)} \\
&= f_{Y|X,Z}(y|x,z) \cdot \frac{f_{X,Z}(x,z)}{f_Z(z)} = f_{Y|X,Z}(y|x,z)f_{X|Z}(x|z)
\end{aligned}$$

Matrix Algebra

A $n \times n$ matrix A is orthogonal if $A^T A = I_n$.
A $n \times n$ matrix A is idempotent if $\forall n \in \mathbb{N}$, $A^n = A$.
Two matrices A_1, A_2 are orthogonal to each other if $A_1 A_2 = 0_n$
If a matrix $Q_{n\times k}$ is idempotent, then $Rank(Q) = tr(Q)$
For any two matrices $A_{n\times k}, B_{k\times l}$, we have $tr(AB) = tr(BA)$, $tr(A + B^T) = tr(A) + tr(B), tr(cA) = c \cdot tr(A), c \in \mathbb{R}$

Asymptotic Properties

- A sequence of random variables X_n converges **in meas squared errors** to a random variable X if $E[(X_n - X)^2] \to 0$
- A sequence of random variables X_n converges **in probability** to a random variable X if $\forall \varepsilon > 0$, $P(|X_n - X| > \varepsilon) \to 0$. We can also denote this as $X_n \xrightarrow{p} X$ or say $X_n - X$ is $o_p(1)$.
- A sequence of random variables X_n converges **in distribution** to a random variable X if $F_{X_n} \to F_X$.
- A sequence of random variables X_n is **bounded in probability** if $\forall \varepsilon > 0$, $\exists b_\varepsilon > 0$, $P(|X_n| \geq b_\varepsilon) \leq \varepsilon$. We denote X_n being bounded in probability as $X_n = O_p(1)$.
- Note that convergence in probability implies convergence in distribution, which implies boundedness in probability.

Continuous Mapping Theorem (CMT)

Let $g(X): supp(X_n) \to supp(X)$. If g is a continuous function on the support of X, we have

$$X_n \xrightarrow{p} X \Rightarrow g(X_n) \xrightarrow{p} g(X)$$
$$X_n \xrightarrow{d} X \Rightarrow g(X_n) \xrightarrow{d} g(X)$$
$$X_n \xrightarrow{p} c \in \mathbb{R} \wedge Y_n \xrightarrow{p} Y \Rightarrow X_n Y_n \xrightarrow{p} cY$$
$$X_n \xrightarrow{p} c \in \mathbb{R} \wedge Y_n \xrightarrow{d} Y \Rightarrow X_n Y_n \xrightarrow{d} cY$$

Stochasic Order Algebra

(i) $o_p(1) + o_p(1) = o_p(1)$
(ii) $o_p(1) + O_p(1) = O_p(1)$
(iii) $o_p(1) \cdot O_p(1) = o_p(1)$
(iv) $(1 + o_p(1))^{-1} = O_p(1)$
(v) $o_p(O_p(1)) = o_p(1)$

Weak Law of Large Numbers (WLLN)

Version 1: If $X_i \sim D_i(X)$ such that $E[X], E[X^2] < \infty$ AND $E[X_i X_j] = 0$, then $\bar{X}_n \xrightarrow{p} E[X]$

Version 2: If $X_i \overset{iid}{\sim} D(X)$ such that $E[X] < \infty$, then $\bar{X}_n \xrightarrow{p} E[X]$

Central Limit Theorem (CLT)

If $X_i \overset{iid}{\sim} D(X)$ such that $E[X_i^2] < \infty$, then

$$\sqrt{n}(\bar{X}_n - \mu) \xrightarrow{d} N(0, Var(X_i))$$

Properties of Estimators

- An estimator $\hat{\theta}$ is biased if $E[\hat{\theta}] \neq \theta$.
- An estimator $\hat{\theta}$ is consistent if $\hat{\theta} \xrightarrow{p} \theta$
- An estimator is asymptotically unbiased if $Bias(\hat{\theta}) \to 0$
- In general, $E[\bar{X}_n] = E[X]$, $Var(\bar{X}_n) = \frac{Var(X)}{n}$
- The best unbiased estimator is an estimator that is unbiased AND has the smallest asymptotic variance
- The best unbiased estimator is an estimator that is unbiased AND has the smallest asymptotic variance, and is a linear function of the observations X_i.
- A sequence of estimators $\hat{\theta}_n$ is $\sqrt{n}-$**consistent** if
$$\sqrt{n}(\hat{\theta}_n - \theta) = O_p(1)$$
- Two sequences of estimators $\hat{\theta}_n, \tilde{\theta}_n$ is $\sqrt{n}-$**asymptotically equivalent** if
$$\sqrt{n}(\hat{\theta}_n - \tilde{\theta}_n) = O_p(1)$$

The Delta Method

For a sequence of estimator $\hat{\theta}_n$ such that

$$\sqrt{n})\hat{\theta}_n - \theta) \xrightarrow{d} N(0, v(\theta))$$

and that $g(x)$ is continuous on $supp(\theta_n)$, we have

$$\sqrt{n}(g(\hat{\theta}_n) - g(\theta)) \xrightarrow{d} N\left(0, g'(\theta)^T \cdot v(\theta) \cdot g'(\theta)\right)$$

Cookbook Approach to Delta Method

Suppose that we want to find the asymptotic distribution of an estimator $\hat{\theta}$:

Step 1: $\hat{\theta} = \bar{X}$?

Yes $\Rightarrow$ CLT, $\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, Var(X))$

No $\Rightarrow$ Step 2

Step 2: Is $\hat{\theta}$ a function of $\bar{X}_n$?

Yes $\Rightarrow$ CLT, $\sqrt{n}(\bar{X} - E[X]) \xrightarrow{d} N(0, Var(X))$

and by Delta method

$\sqrt{n}(g(\bar{X}) - g(E[X])) \xrightarrow{d} N(0, g'(\theta)Var(X)g(\theta))$

No $\Rightarrow$ Step 3

Step 3: Is $\hat{\theta}$ a function of some $\bar{Y}_n$? Most likely yes, $\Rightarrow$ CLT, $\sqrt{n}(\bar{Y} - E[Y]) \xrightarrow{d} N(0, Var(Y))$

and by Delta method

$\sqrt{n}(g(\bar{Y}) - g(E[Y])) \xrightarrow{d} N(0, g'(\theta)Var(Y)g(\theta))$

No, then we likely cannot use the delta method.

Common Estimators

Method of Moments: Figure out which moment you want to estimate, then use the sample analogue as the estimator. (See Example).

Maximum Likelihood Estimators: Figure out the joint (log-)likelihood function of the $n-$sample, check first and second order conditions so that you have an estimator that maximizes the joint likelihood function. (Note that MLE are usually consistent asymptotically most efficient, but they are often biased.)

Common MLE for First moments:

Poisson	$\bar{X}_n$
Exponential	$\bar{X}_n$
Normal	$\bar{X}_n$
Binomial/Bernoulli	$\bar{X}_{nm}/\bar{X}_n$
Uniform	$\max \lvert X_i \rvert$ or $\max X_i$

Common Estimators for Variance:

Normal	$\hat{\sigma}^2_{MLE} = \frac{1}{n}\sum(X_i - \bar{X})^2$
General	$\hat{\sigma}^2_{Unbiased} = S^2 = \frac{1}{n-1}\sum(X_i - \bar{X})^2$

Cramer-Rao Lower Bound

For any distribution that satisfies:

1. $f(x;\theta)$ has bounded support in x and the bounds do not depend on θ (so CRLB does not work on uniform)
2. $f(x;\theta)$ has infinite support, is continuously differentiable, and integrable for all θ

The lower bound of the asymptotic variance of **any estimator** for parameters of the distribution can be defined as

$$V(\hat{\theta}) \geq \frac{1}{I(\theta)}$$

where $I(\theta)$ is the Fisher information matrix defined as:

$$I(\theta) = nE\left[\left(\frac{\partial l(X;\theta)}{\partial \theta}\right)^2\right] = -nE\left[\frac{\partial^2 l(X;\theta)}{\partial \theta^2}\right]$$

where $l(X;\theta)$ is the log-likelihood function for a single observation.

Notice that, by construction, $\hat{\theta}_{MLE}$ always achieves CRLB, but it is also almost always biased.

Hypothesis Testing

When we observe data, we create a test statistics T_n. Putting T_n against a critical region C_α given the pre-determined *size* of the test α. We reject the null hypothesis if $T_n \in C_\alpha$

- Size $\alpha = P(Reject \mid H_0) = P(T_n \in C_n \mid H_0$
- Power $1 - \beta = \beta(\theta) = P(Reject \mid H_0)$

For a test, we want to maximize power ($1-\beta$ or equivalently $\beta(\theta)$) given a specific α.

p-value is the minimal α needed to reject H_0 with the data observed. One should not think about this as a probability. Formally,

$$p - value \equiv \inf\{\alpha \in (0,1) \mid T_n \in C_\alpha\}$$

This definition is important because it gives clear guidelines on calculating p-values for a discrete R.V.

A sequence of tests is **consistent** if it has asymptotic power $\lim_{n\to\infty} \beta(\theta) = \lim_{n\to\infty} P(reject\, H_0 \mid \theta) = 1, \forall \theta \in \Theta_1$

Local Power Analysis

Consider, instead, a sequence of alternative hypotheses $H_{n1}: \mu_n = \mu_0 + \frac{\delta}{\sqrt{n}}$ against the $H_0: \mu = \mu_0$. Then we can calculate the power of a sequence of tests $\beta(\theta) = P(reject\, H_0 \mid H_{n1})$. As $n \to \infty$, H_1 gets closer and closer to H_0. This gives us an idea of what the local power looks like when designing a test, **not designing an estimator**.

Confidence Intervals

By CLT, if we have iid samples with finite second moment, we know that we can achieve some type of asymptotic normality of our estimators. For example, say

$$\sqrt{n}(\hat{\theta} - \theta) \xrightarrow{d} N(0, v(\theta))$$

then we can also write

$$\sqrt{n}\left(\frac{\hat{\theta} - \theta}{\sqrt{v(\theta)}}\right) \xrightarrow{d} N(0,1)$$

This means that we can use the $z-$table to form the interval F_α for evidence that we would **fail to reject under** H_0 as:

$$F_\alpha = (\mu_0 - z_\alpha se(\hat{\theta}))$$

such that $T_n \in F_\alpha \Rightarrow$ Fail to reject H_0.

Multivariate Normal Distribution

Conditional Normal

Consider random vectors $X_{m\times 1}, Y_{n\times 1}$ that are jointly normally distributed:

$$\begin{pmatrix} X \\ Y \end{pmatrix} \sim Normal\left(\begin{pmatrix} \mu_X \\ \mu_y \end{pmatrix}, \begin{pmatrix} \Sigma_{XX} & \Sigma_{XY} \\ \Sigma_{YX} & \Sigma_{YY} \end{pmatrix}\right)$$

where

$$\Sigma_{XY} = Cov(X,Y)_{m\times n} = \sum_{YX}'$$

Then,

$$\begin{aligned} Y|X &\sim Normal(\alpha + B'X, \Sigma_{Y|X}) \\ B &= \Sigma_{XX}^{-1}\Sigma_{XY} \\ \alpha &= \mu_Y - B'\mu_X \\ \Sigma_{Y|X} &= \Sigma_{YY} - \Sigma_{YX}\Sigma_{XX}^{-1}\Sigma_{XY} \end{aligned}$$

Diagonalization of the Variance Matrix

A real, symmetric matrix Σ (which we assume variance matrices are), $\Sigma = QDQ'$ where Q is an orthonormal matrix ($QQ' = Q'Q = I$) and D is a diagonal matrix of eigenvalues. If we further assume that A is **positive definite**, then we can define $\Sigma^{-\frac{1}{2}} = QD^{-\frac{1}{2}}Q'$ where λ_i's are the eigenvalues and Q is made of corresponding eigenvectors.

$$D^{-\frac{1}{2}} = \begin{pmatrix} \lambda_1^{-\frac{1}{2}} & 0 & \cdots & 0 \\ 0 & \lambda_2^{-\frac{1}{2}} & \cdots & 0 \\ \vdots & & \ddots & 0 \\ 0 & \cdots & & \lambda_n^{-\frac{1}{2}} \end{pmatrix}$$

If matrix B is symmetric and idempotent ($B^n = B$), then $X'BX = X'B'BX = (BX)'BX$.

If matrix $B_{n\times n}$ is symmetric, idempotent, and real with rank m ($\leq n$), it is diagonlizable with $B = QDQ'$ where D is a diagonal matrix with a total of m 1's in the diagonal.

$X \sum N(0, I_n) \Rightarrow X'_{1\times n}B_{n\times n}X_{n\times 1} \sim \chi^2_m$

Selected Problems

Find $f_Y(y)$ where $Y = e^X$ and $f_X = \frac{1}{\sigma^2}x \cdot exp(-\frac{x^2}{2\sigma^2})$

Sol: Since e^X is strictly monotonic, we can use the formula

$$f_Y(y) = [|\frac{dx(y)}{dy}|]f_x(g^{-1}(y)) =$$
$$\frac{1}{y}\frac{1}{\sigma^2}ln(y)e^{-(ln(y)/\sigma)^2/2},\ y \in (1, \infty)$$

Find $f_Y(y)$ where $Y = \frac{4}{3}X - X^2$ and $X \sim Uniform[0,1]$

Sol:

$$\begin{aligned}F_Y(y) &= P(\frac{4}{3}X - X^2 \le y) = P((X - \frac{2}{3})^2 \ge \frac{4}{9} - y)\\ &= 1 - P((X - \frac{2}{3})^2 \le \frac{4}{9} - y)\\ &= 1 - P(\frac{2}{3} - (\frac{4}{9} - y)^{\frac{1}{2}} \le X \le \frac{2}{3} + (\frac{4}{9} - y)^{\frac{1}{2}})\\ &= 1 - [F_X(\frac{2}{3} + (\frac{4}{9} - y)^{\frac{1}{2}}) - F_X(\frac{2}{3} - (\frac{4}{9} - y)^{\frac{1}{2}}]\end{aligned}$$

Notice that at $y \le \frac{3}{9} = \frac{1}{3}$, $F_X(\frac{2}{3} + (\frac{4}{9} - y)^{\frac{1}{2}}) = 1$ since $x \in [0,1]$. Hence we have the CDF:

$$F_Y(y) = \begin{cases} \frac{2}{3} - (\frac{4}{9} - y)^{\frac{1}{2}} & , y \le \frac{1}{3} \\ 1 - 2(\frac{4}{9} - y)^{\frac{1}{2}} & , \frac{1}{3} < y \le \frac{4}{9} \\ 0 & , \text{otherwise} \end{cases}$$

and hence we have the PDF of Y as:

$$f_Y(y) = \begin{cases} \frac{1}{2}(\frac{4}{9} - y)^{-\frac{1}{2}} & , y \le \frac{1}{3} \\ (\frac{4}{9} - y)^{-\frac{1}{2}} & , \frac{1}{3} < y \le \frac{4}{9} \\ 0 & , \text{otherwise} \end{cases}$$

$X \sim Gamma(\alpha, \beta)$, show that $P(X \ge 2\alpha\beta) \le (2/e)^\alpha$.

Sol: Using Markov's Inequality, we can bound the probability by:

$$\begin{aligned}P(X \ge 2\alpha\beta) &= P(e^{tX} \ge e^{t2\alpha\beta}) \le \frac{E[e^{tX}]}{e^{t2\alpha\beta}}\\ &= \underbrace{\frac{(1-\beta t)^{-\alpha}}{e^{t2\alpha\beta}}}_{\text{using } t = \frac{1}{2\beta} < \frac{1}{\beta}} = \frac{(\frac{1}{2})^{-\alpha}}{e^\alpha} = (\frac{2}{e})^\alpha\end{aligned}$$

$X \sim Normal(\mu, \sigma^2)$, show $E[|X - \mu|] = \sigma\sqrt{2/\pi}$

Sol: Notice that $N(\mu, \sigma^2)$ is symmetric about $x = \mu$, so

$$\begin{aligned}E[|X - \mu|] &= 2 \cdot \int_\mu^\infty \frac{x-\mu}{\sigma\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}dx\\ &= 2(-\frac{2\sigma}{2\sqrt{2\pi}}e^{-\frac{(x-\mu)^2}{2\sigma^2}}|_\mu^\infty = 2(0 - (-\frac{\sigma}{\sqrt{2\pi}}e^0))\\ &= \frac{2\sigma}{\sqrt{2\pi}} = \sigma\frac{\sqrt{2}}{\sqrt{\pi}} = \sigma\sqrt{\frac{2}{\pi}}\end{aligned}$$

Find the moment generating function for $f(x) = \frac{1}{4}exp\left(-\frac{|x-a|}{2}\right)$, $x, \alpha \in \mathbb{R}$

Sol:

$$\begin{aligned}\psi_X(t) &= \int_{-\infty}^\infty \frac{1}{4}e^{-\frac{|x-\alpha|}{2}}e^{tx}dx\\ &= \int_{-\infty}^\alpha \frac{1}{4}e^{\frac{x-\alpha}{2}}e^{tx}dx + \int_\alpha^\infty \frac{1}{4}e^{\frac{\alpha-x}{2}}e^{tx}dx\\ &= \frac{1}{4}e^{-\frac{\alpha}{2}}\int_{-\infty}^\alpha e^{\frac{2t+1}{2}x}dx + \frac{1}{4}e^{\frac{\alpha}{2}}\int_\alpha^\infty e^{\frac{2t-1}{2}x}dx\\ &= \frac{1}{4}e^{-\frac{\alpha}{2}}\frac{2}{2t+1}e^{\frac{2t+1}{2}x}|_{-\infty}^\alpha + \frac{1}{4}e^{\frac{\alpha}{2}}\frac{2}{2t-1}e^{\frac{2t-1}{2}x}|_\alpha^\infty\\ &= \frac{2}{4(2t+1)}e^{\alpha t} + \frac{2}{4(2t-1)}e^{\alpha t}\\ &= \frac{2}{(2t+1)(2t-1)}e^{\alpha t}\end{aligned}$$

Find the moment generating function for $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$

Since X is a discrete random variable following the Poisson(λ) distribution, its MGF is:

$$\begin{aligned}\psi(t) &= \sum_{x=0}^\infty \frac{\lambda^x e^{-\lambda}}{x!}e^{tx} = e^{-\lambda}\sum_{x=0}^\infty \frac{\lambda^x e^{tx}}{x!}\\ &= e^{-\lambda}\sum_{x=0}^\infty \frac{(\lambda e^t)^x}{x!}\\ &= \underbrace{e^{-\lambda}e^{\lambda e^t}}_{\text{Using the power series expansion for exponential function}}\\ &= e^{\lambda(e^t - 1)}\end{aligned}$$

Find the moment generating function for $P(X = x) = \frac{e^{-\lambda}\lambda^x}{x!}$, $x \in \mathbb{N} \cup \{0\}$, $\lambda > 0$

Sol: Since X is a discrete random variable following the Poisson(λ) distribution, its MGF is:

$$\begin{aligned}\psi(t) &= \sum_{x=0}^\infty \frac{\lambda^x e^{-\lambda}}{x!}e^{tx} = e^{-\lambda}\sum_{x=0}^\infty \frac{\lambda^x e^{tx}}{x!} = e^{-\lambda}\sum_{x=0}^\infty \frac{(\lambda e^t)^x}{x!}\\ &= \underbrace{e^{-\lambda}e^{\lambda e^t}}_{\text{Using the power series expansion for exponential function}}\\ &= e^{\lambda(e^t - 1)}\end{aligned}$$

($\star$) Suppose in a shop on average ten customers come in per hour. What is the probability when you enter that you would have to wait more than twenty minutes for the next customer to come in?

Sol: The number of minutes we, on average, have to wait follows the **continuous** distribution $exponential(\frac{1}{6})$, so

$$\begin{aligned}P(X \ge 20) &= 1 - P(X \le 20)\\ &= 1 - (1 - e^{-\frac{1}{6}\cdot 20}) = \frac{1}{e^{\frac{10}{3}}} = 0.0357\end{aligned}$$

Notice that there are several ways to specify the distribution for this problem. The following specifications are equivalent:

$$H \sim Gamma(1, \beta) \qquad \beta = \frac{1}{\lambda} = \frac{1}{10} \tag{1}$$
$$M \sim Gamma(1, \beta) \qquad \beta = \frac{1}{\lambda} = 6 \tag{2}$$
$$H \sim Exponential(\lambda) \qquad \lambda = \frac{1}{10} \tag{3}$$
$$M \sim Exponential(\lambda) \qquad \lambda = 6 \tag{4}$$

where H is the random variable representing the hours before an event and M is the random variable representing the minutes before an event. For each case, the heuristic description of the distribution is:

H is distributed such that per hour ($\alpha = 1$), there are 10 customers ($\lambda = \frac{1}{\beta} = 10$).

M is distributed such that per minute($\alpha = 1$), there are $\frac{1}{6}$ customers ($\lambda = \frac{1}{\beta} = \frac{1}{6}$).

$$f_{Y|X}(y|x) = \frac{2y+4x}{1+4x}$$
$$f_X(x) = \frac{1+4x}{3}$$

Find $f_{X|Y}$. **Sol:**

$$f_{X|y} = \frac{f_{Y|x} \cdot f_X}{f_Y} = \frac{f_{Y|x} \cdot f_X}{\int_0^1 f_{Y|x} \cdot f_X dx}$$
$$= \frac{f_{Y|x} \cdot f_X}{\int_0^1 \frac{2y+4x}{3} dx} = \frac{f_{Y|x} \cdot f_X}{\int_0^1 \frac{2y+4x}{3} dx} = \frac{\frac{2y+4x}{3}}{\frac{2y+2}{3}}$$
$$= \frac{y+2x}{y+1}, \; 0 < x < 1, 0 < y < 1$$

conditional density is 0 otherwise

$$
\begin{aligned}
Y &= X + Z - 2XZ + U \\
E[U|X,Z] &= 0 \\
E[Z|X] &= 3 + 4X
\end{aligned}
$$

Find $E[Y|X,Z]$ and $E[Y|X]$. **Sol:**

$$
\begin{aligned}
E[Y|X,Z] &= E[X + Z - 2X \cdot Z + U|X,Z] \\
&= X + Z - 2X \cdot Z \\
E[Y|X] &= E[E[Y|X,Z]|X] = E[X + Z - 2X \cdot Z|X] \\
&= X + E[Z|X] - 2XE[Z|X] \\
&= X + (3 + 4X) - 2X(3 + 4X) \\
&= -8X^2 - X + 3
\end{aligned}
$$

Let $X_i \overset{iid}{\sim} D(X)$ such that $f(x;\theta) = \theta x^{\theta-1}$, $supp(X) = (0,1), \theta \in R_{++}$.

Derive $\hat{\boldsymbol{\theta}}_{\boldsymbol{MLE}}$.

$$L(n;\theta) = \prod_{i=1}^{n} \theta X_i^{\theta-1} = \theta^n \prod_{i=1}^{n} X_i^{\theta-1}$$
$$l(n;\theta) = nln(\theta) + \sum_{i=1}^{n} (\theta-1) ln(X_i)$$
$$s(n;\theta) = \frac{n}{\theta} + \sum_{i=1}^{n} ln(X_i) = 0$$
$$\frac{\partial s(n;\theta)}{\partial \theta} = -\frac{n}{\theta^2} < 0$$

So the MLE estimator is

$$\hat{\theta}_{MLE} = -\frac{n}{\sum\limits_{i=1}^{n} ln(X_i)}$$

Let $Y_i = ln(X_i)$

$$f_Y = e^y \theta (e^y)^{\theta-1} = \theta e^{y\theta}$$
$$E[Y] = \int_{-\infty}^{0} y \cdot \theta e^{y\theta} dy$$
$$= y e^{y\theta}|_{-\infty}^{0} - \frac{1}{\theta} e^{y\theta}|_{-\infty}^{0} = -\frac{1}{\theta}$$
$$E[Y^2] = \int_{-\infty}^{0} y^2 \cdot \theta e^{y\theta} dy$$
$$= y^2 e^{y\theta}|_{-\infty}^{0} - 2y\frac{1}{\theta} e^{y\theta}|_{-\infty}^{0} + \frac{2}{\theta^2} e^{y\theta}|_{-\infty}^{0}$$
$$= \frac{2}{\theta^2}$$
$$Var(Y) = \frac{2}{\theta^2} - \frac{1}{\theta^2} = \frac{1}{\theta^2}$$

By CLT,

$$\sqrt{n}\left(\frac{\sum\limits_{i=1}^{n} ln(X_i)}{n} + \frac{1}{\theta}\right) \xrightarrow{d} N(0, \frac{1}{\theta^2})$$

So by the delta method,

$$\sqrt{n}\left(-\frac{n}{\sum\limits_{i=1}^{n} ln(X_i)} - \theta\right) \xrightarrow{d} N(0, \theta^2)$$

and hence

$$\hat{\theta}_{MLE} = -\frac{n}{\sum\limits_{i=1}^{n} ln(X_i)} \overset{a}{\sim} N(\theta, \frac{\theta^2}{n})$$

Derive $\tilde{\boldsymbol{\theta}}_{\boldsymbol{MOM}}$.

Since $E[X_i] = \frac{\theta}{\theta+1}$, MOM estimator can be derived from

$$\bar{X} = \frac{\hat{\theta}}{\hat{\theta}+1}$$
$$\Rightarrow \tilde{\theta}_{MOM} = \frac{\bar{X}}{1-\bar{X}}$$

We know that, by CLT,

$$\sqrt{n}(\bar{X} - \frac{\theta}{\theta+1}) \xrightarrow{d} N\left(0, \frac{\theta}{(\theta+2)(\theta+1)^2}\right)$$

So by the delta method

$$\sqrt{n}(\tilde{\theta}_{MOM} - \theta) \xrightarrow{d} N\left(0, \frac{\theta(\theta+1)^2}{\theta+2}\right)$$
$$\tilde{\theta}_{MOM} \overset{a}{\sim} N\left(\theta, \frac{\theta(\theta+1)^2}{n(\theta+2)}\right)$$

Compare the asymptotic variance of the two estimators.

$$AVar(\hat{\theta}_{MLE}) - AVar(\tilde{\theta}_{MOM})$$
$$= \frac{\theta}{n}\left(\frac{\theta(\theta+2)}{\theta+2} - \frac{(\theta+1)^2}{\theta+2}\right)$$
$$= -\frac{\theta}{n(\theta+2)} < 0$$

The MLE estimator is more efficient than the MOM estimator.