ST4231 Cheatsheet

Ng Xuan Jun

Useful Distributions

Bernoulli Random Variable: The random variable only has 2 possible outcomes. Probability of one of them is p

$$X \sim Bernoulli(p)$$

Probability Mass Function (PMF):

$$P(X=k)=\begin{cases} p, & k=1 \\ 1-p, & k=0 \end{cases}$$

Expectation: $E(X)=p$

Variance: $Var(x)=p(1-p)$

Indicator Function is a Bernoulli Random Variable

$$1_A=\begin{cases} 1, & if\ A\ happens \\ 0, & if\ A\ doesnt\ happen \end{cases}$$

Binomial Random Variable: Number of successes in n Bernoulli trials

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

Probability Mass Function (PMF):

$$P(X=k)=\binom{n}{k}p^k(1-p)^{n-k}, \qquad k=0,1,\cdots,n$$

Expectation: $E(X)=np$

Variance: $Var(X)=np(1-p)$

If $X_1,\cdots,X_n$ are i.i.d. with common distribution $Bernoulli(p)$, then $X_1+\cdots+X_n \sim Bin(n,p)$

Geometric Random Variable: Number of Bernoulli trials to obtain the **first** success

$$X \sim Geometric(p)$$

Probability Mass Function (PMF):

$$P(X=k)=p(1-p)^{k-1}, \qquad k=1,2,3,\cdots$$

Expectation: $E(X)=\frac{1}{p}$

Variance: $Var(X)=\frac{1-p}{p^2}$

Poisson Random Variable: The number of events occurring in a fixed time interval or region of opportunity. Number of events per single unit of time

$$X \sim Poi(\lambda)$$

Probability Mass Function (PMF):

Note that $\lambda > 0$

$$P(X=k)=\frac{\lambda^k}{k!}e^{-\lambda}, \qquad k=0,1,2,\cdots$$

Expectation: $E(X)=\lambda$

Variance: $Var(X)=\lambda$

Poisson Approximation: When n is large and p is small, np is moderate

$$Bin(n,\ p) \to Poisson\ (np)$$

Uniform Random Variable:

$$X \sim Uniform(a,b)$$

Probability Density Function (PDF):

$$f(x)=\begin{cases} \frac{1}{b-a}, & a \le x \le b \\ 0, & otherwise \end{cases}$$

Cumulative Distribution Function (CDF):

$$F(x)=\begin{cases} 0, & x<a \\ \frac{x-a}{b-a}, & a \le x < b \\ 1, & b \le x \end{cases}$$

Expectation: $E(X)=\frac{a+b}{2}$

Variance: $Var(X)=\frac{(b-a)^2}{12}$

Standard Uniform Distribution:

$$X \sim Uniform(a,b)$$

$$U(0,\ 1) \equiv Beta(1,\ 1)$$

PDF: $f(x)=1,\ \ for\ 0 \le x \le 1$

CDF: $F(x)=x,\ \ for\ 0 \le x \le 1$

Expectation: $E(X)=\frac{1}{2}$

Variance: $Var(X)=\frac{1}{12}$

Transform to Uniform(a, b):

$$Y=(b-a)X+a$$

Normal Random Variable:

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

Probability Density Function (PDF):

$$f(x)=\frac{1}{\sqrt{2\pi}\sigma}\exp\left(-\frac{(x-\mu)^2}{2\sigma^2}\right), \quad -\infty<x<\infty$$

Cumulative Distribution Function (CDF):

$$F(x)=\int_{-\infty}^{\infty} f(x)dx, \quad -\infty<x<\infty$$

Expectation: $E(X)=\mu$

Variance: $Var(X)=\sigma^2$

Standard Normal Distribution: $N\ (0,\ 1)$

If $Z \sim N(0,\ 1)$ then $\mu+\sigma Z \sim N(\mu,\sigma^2)$

Probability Density Function (PDF):

$$f(x)=\frac{1}{\sqrt{2\pi}}\exp\left(-\frac{1}{2}x^2\right), \quad -\infty<x<\infty$$

Cumulative Distribution Function (CDF):

$$F(x)=\int_{-\infty}^{\infty} f(x)dx, \quad -\infty<x<\infty$$

Expectation: $E(X)=0$

Variance: $Var(X)=1$

d-dimensional normal with mean μ and covariance matrix Σ

$$f(x;\mu,\Sigma)=\frac{1}{(2\pi)^{\frac{d}{2}}|\Sigma|^{\frac{1}{2}}}\exp\left(-\frac{1}{2}(x-\mu)^T\Sigma^{-1}(x-\mu)\right), \quad -\infty<x<\infty$$

Note that $|\Sigma|$ is the determinant of Σ

Exponential Random Variable:

Note that $\lambda > 0$

$$X \sim Exp(\lambda)$$

Probability Density Function (PDF):

$$F(x)=\begin{cases} \lambda e^{-\lambda x}, & x>0 \\ 0, & x \le 0 \end{cases}$$

Cumulative Distribution Function (CDF):

$$F(x)=\begin{cases} 1-e^{-\lambda x}, & x>0 \\ 0, & x \le 0 \end{cases}$$

Expectation: $E(X)=\frac{1}{\lambda}$

Variance: $Var(X)=\frac{1}{\lambda^2}$

Memoryless Property: For any $X \sim Exp(\lambda)$

$$P(X>s+t|\ X>s)\ =P(X>t)$$

Gamma Random Variable:

Note that shape parameter $a>0$, rate parameter $b>0$

$$X \sim Gamma(a,b)$$

Probability Density Function (PDF):

$$g(x)=\begin{cases} \frac{\lambda^a}{\Gamma(a)}x^{a-1}e^{-\lambda t}, & x \ge 0 \\ 0, & t<0 \end{cases}$$

Cumulative Distribution Function (CDF):

$$G(x)=\frac{1}{\Gamma(a)}\gamma(\alpha,\ \lambda X)$$

Expectation: $E(X)=\frac{\alpha}{\lambda}$

Variance: $Var(X)=\frac{\alpha}{\lambda^2}$

Gamma Function: $\Gamma(z)=\int_0^\infty t^{z-1}e^{-t}dt$

Recursion Property: $\Gamma(a+1)=a\Gamma(a)$

Gamma Function Computation: $\Gamma(x)=(x-1)!$

Sum of Gamma Random Variables: Let $X_1,\cdots,X_k$ be independent random variables, assume $X_i \sim Gamma(a_i,b)$ for each i. Then

$$X_1+\cdots+X_k \sim Gamma\ (a_1+\cdots+a_k,b)$$

Connection with Standard Normal: If $Z \sim N(0,1)$ then $Z^2 \sim Gamma\left(\frac{1}{2},\frac{1}{2}\right) \sim \chi^2(1)$

Connection with Chi Squared: Assume $Z_1,\cdots,Z_k$ are i.i.d N(0, 1) random variables. Then

$$Z_1^2+\cdots+Z_k^2 \sim Gamma\left(\frac{k}{2},\frac{1}{2}\right) \sim \chi^2(k)$$

Connection with Exponential Distribution: If $X_1,\cdots,X_n$ i.i.d $Exp(\lambda)=Gamma(1,\lambda)$

$$X_1+\cdots+X_n \sim Gamma(n,\lambda)$$

Scaling: If $X \sim Gamma(a,b)$ then $\lambda X \sim Gamma\left(a,\frac{b}{\lambda}\right)$

Beta Random Variable:

Note that $a>0,\ b>0$

$$X \sim Beta(a,b)$$

Probability Density Function (PDF):

$$f(x)=\frac{\Gamma(\alpha+\beta)}{\Gamma(\alpha)\Gamma(\beta)}x^{a-1}(1-x)^{b-1}, \ \ 0 \le x \le 1$$

Expectation: $E(X)=\frac{a}{a+b}$

Variance: $Var(X)=\frac{ab}{(a+b)^2(a+b+1)}$

Swap of parameters: If $X \sim Beta(a,b)$, then $1-X \sim Beta(b,a)$

If $X \sim Gamma(a,\beta),\ Y \sim Gamma(b,\beta)$ and X,Y are independent, then

$$\frac{X}{X+Y} \sim Beta(a,b)$$

Order Statistics: If $X_1,\cdots,X_n$ are i.i.d from Uniform(0,1) and $X_{(1)} \le \cdots \le X_{(n)}$ are their order statistics, then for $k=1,\cdots,n$

$$X_{(k)} \sim Beta(k,n+1-k)$$

- Useful to know when we want to generate Beta distribution, we can just draw iid uniform and order them then pick the kth one

Cauchy Random Variable

$$X \sim Cauchy(x_0,\gamma)$$

Probability Density Function (PDF):

$$f(x)=\frac{1}{\pi\gamma}\left(\frac{1}{1+\left(\frac{x-x_0}{\gamma}\right)^2}\right), \qquad -\infty<x<\infty$$

Cumulative Distribution Function (CDF):

$$F(x)=\frac{1}{\pi}\arctan\left(\frac{x-x_0}{\gamma}\right)+\frac{1}{2}, \qquad -\infty<x<\infty$$

Inverse CDF:

$$F^{-1}(u)=\gamma\tan[\pi(u-0.5)]+x_0, \qquad u \in [0,1]$$

Standard Cauchy Random Variable:

$$X \sim Cauchy(0,1)$$

Probability Density Function (PDF):

$$f(x)=\frac{1}{\pi}\left(\frac{1}{1+x^2}\right), \qquad -\infty<x<\infty$$

Cumulative Distribution Function (CDF):

$$F(x)=\frac{1}{\pi}\arctan x+\frac{1}{2}, \qquad -\infty<x<\infty$$

Inverse CDF:

$$F^{-1}(u)=\tan[\pi(u-0.5)], \qquad u \in [0,1]$$

Scaling and Shifting of Random Variables: Suppose that X is a continuous random variable with pdf $f(x)$

- **Shift**: If a is a real number, then pdf of $X+a$ is $f(x-a)$
- **Scale**: If b is a positive number, then the pdf of bX is $b^{-1}f\left(\frac{x}{b}\right)$

Let $X_1,X_2,\cdots$ be a sequence of iid random variables with mean μ and variance σ^2. We define the n-th sample mean and sample variance by:

Sample Mean: The mean of the sample that we are currently looking at

$$\bar{X}_n=\frac{1}{n}\sum_{i=1}^n X_i, \qquad E(\bar{X}_n)=\mu$$

Sample Variance: Variance of the sample data that we are currently looking at

$$S_n^2=\frac{1}{n-1}\sum_{i=1}^n(X_i-\bar{X})^2, \qquad E(S_n^2)=\sigma^2$$

Variance of Sample Mean: Variation of the sample means that we will get over the n samples

$$Var(\bar{X}_n)=\frac{\sigma^2}{n}$$

Change of variable formula: Suppose U and V are functions of X and Y, $u=g_1(x,\ y)\ \ v=g_2(x,\ y), J(x,\ y) \ne 0$

Multivariable Joint Density of U and V:

$$f_{UV}(u,\ v)=f_{XY}\big(h_1(u,\ v),\ h_2(u,\ v)\big)|J(x,\ y)|$$

Note that $h_1(u,\ v)$ is x represented by u, v only. $h_2(u,\ v)$ is y represented by u, v only.

Jacobian: $J(x,\ y)=det\begin{bmatrix} \frac{\partial u}{\partial x} & \frac{\partial u}{\partial y} \\ \frac{\partial v}{\partial x} & \frac{\partial v}{\partial y} \end{bmatrix}=\left(\frac{\partial u}{\partial x}\right)\left(\frac{\partial v}{\partial y}\right)-\left(\frac{\partial v}{\partial x}\right)\left(\frac{\partial u}{\partial y}\right)$, Rows – Functions, Columns, Variables

Single Variable: Suppose $g(x)$ is a one-to-one differentiable function. If X has pdf $f_X(x)$ and $Y=g(X)$ then pdf of Y is:

$$f_Y(y)=f_X\big(g^{-1}(y)\big)\cdot\left|\frac{dg^{-1}(y)}{dy}\right|$$

$g^{-1}(y)$ is just x in terms of y and we substitute it into f_x

Note that if $g(x,y)$ is not one-to-one, we break it into intervals such that it is one to one and we just add up the distribution on the range where they are one-to-one

(Strong) Law of Large Numbers (SLLN): Suppose that the random variable X has finite first moment(i.e. $E[|X|]<\infty$), then the sample mean (based on a random sample $\{X_1,\cdots,X_n\}$) converges **almost surely** to the population mean

$$\lim_{n\to\infty}\frac{(X_1+\cdots+X_n)}{n}=E[X]$$

Central Limit Theorem (CLT iid version): Suppose that the random variable X has finite second moment (i.e. $E[X^2]<\infty$), then the following **convergence in distribution** holds

$$\lim_{n\to\infty}\sqrt{n}(\bar{X}_n-\mu)=N(0,\sigma^2)$$

- $\bar{X}_n-\mu$ converges to 0 in order of $n^{-\frac{1}{2}}$
- In the multivariate case, replace σ^2 by the covariance matrix

Monte Carlo Integration:

- Used for estimating some kind of parameter

Consider the density function to be $f(u)=1, U \sim Uniform(0,1)$

$$\theta=\int_0^1 g(U)\cdot 1\ du=E_U[g(U)]$$

By SLLN, if $\int_0^1|g(x)|\ dx<\infty$, then with probability 1

$$\frac{1}{k}\sum_{i=1}^k g(U_i) \to E[g(U)]=\theta\ \ as\ k\to\infty$$

If we can generate large number of random numbers from Uniform(0, 1) then we can approximate θ by the average of the $g(u_i)$

$$\hat{\theta}=\frac{1}{k}\sum_{i=1}^k g(u_i)$$

Expectation is an integral and probability is an expectation of an indicator function

$$\theta=E\big(g(X,Y)\big)=\int_S 1_{condition}(x,y)\,f(x,y)dx \approx \frac{1}{n}\sum_{i=1}^n 1_{condition}(x,y) \xrightarrow{SLLN} \theta$$

Discrete Random Variable Generation:
Compute the probability for each of the possible values for the pmf. Then we just generate a uniform distribution to check which of the probability range it lies within.
Algorithm:
1. Generate $U \sim Uniform(0,1)$
2. If $U < p_0$, set $X = x_0$ and stop
3. If $U < p_0 + p_1$, set $X = x_1$ and stop
4. $\vdots$
5. Otherwise, set $X = x_n$

Inversion Method (Continuous Random Variable):
For a given random variable X, if we want to generate it, we can do the following. If we are given the pdf of the random variable X, $f(x)$:
1. Integrate $f(x)$ over the entire range to get the CDF, $F(x)$
2. Let $U \sim Uniform(0,1)$. Set $U = F(x)$ and find the inverse of the cdf $F^{-1}(u) = X$
3. Once we have found the inverse CDF, we can just generate a uniform distribution and put inside the inverse CDF to get one X

Algorithm:
1. Generate $U \sim Uniform(0,1)$
2. Set $X = F^{-1}(U)$

Note:
1. For $Exp(\lambda)$, we can use the following to generate the random variable:
$$X = -\frac{1}{\lambda}\log U$$
2. For Gamma(n, λ), we can use the following to generate the random variable:
Note that we are making use of the fact that Gamma(n,λ) is the sum of n Exp(λ)
$$X = -\frac{1}{\lambda}\log(U_1 \cdots U_n)$$

Fundamental Theorem of Simulation
If X is a random variable with pdf $f(x)$, then simulating X is equivalent to simulating a pair fo random variables (X, U) jointly from
- Basically we can just sample some value of x and check that the probability of randomly getting that value of x is within the density of that particular x. If we can get the values of X like that, then it will have a pdf of $f(x)$
$$(X,U) \sim Uniform\{(x,u): 0 < u < f(x)\}$$

Rejection Sampling:
Quick ways to check that $\sup_x \frac{f(x)}{g(x)} < +\infty$ (But still need to rigorously show, this can give a brief idea)
1. Domain of $g(x)$ should cover the domain of $f(x)$
2. The tails of the proposal $g(x)$ should be heavier than the tails of $f(x)$

Rigorous ways to check:
1. Differentiate the ratio of $\frac{f(x)}{g(x)}$ and find the maximum value that the ratio can attain
2. Try to observe what will happen to the ratio $\frac{f(x)}{g(x)}$ when $x \to +\infty$. Look at what kind of function it will look like and make the conclusion from there

Theoretical number of simulations required to get 1 acceptance: $M = \sup_x \frac{f(x)}{g(x)}$

Logical steps to do (When we are computing):
1. Try to imagine the shape of $g(x)$ and $f(x)$, when one increases, the other should increase also, vice versa
2. Find the value of $M = \sup_x \frac{f(x)}{g(x)}$. Check that it exists and state the value where we can compute the maximum value using the rigorous way to check
3. Specify the rejection function of $\frac{f(x)}{Mg(x)}$ and our U needs to be within the rejection function range else it will be rejected
4. Generate Y using some kind of method (normally inversion)

Algorithm:
1. Generate $Y \sim g$
2. Generate $U \sim Uniform(0,1)$
3. If $U \le \frac{f(Y)}{Mg(Y)}$, then accept: set $X = Y$ and stop. Otherwise, reject and return to step 1

Unknown Normalising Constant:
If we only know $f(x)$ up till a certain normalising constant, it will work the same, just take the ratio and supremum to be:
$$\frac{\tilde{f}(x)}{g(x)}, \qquad \sup_x \frac{\tilde{f}(x)}{\tilde{M}g(x)}$$

Polar Method for Bivariate Normal:
$$S = R^2 = X^2 + Y^2, \quad \tan\theta = \frac{Y}{X}, \quad X = R\cos(\theta), \quad Y = R\sin(\theta)$$
Change of variable from (X,Y) to (S,θ)
$$f(s,\theta) = \frac{1}{2}e^{-\frac{s}{2}}\frac{1}{2\pi}, \quad 0 < s < \infty, 0 < \theta < 2\pi$$
$$S = R^2 \sim Exp\left(\frac{1}{2}\right) \text{ and } \theta \sim Uniform(0, 2\pi)$$

Box-Muller Algorithm v1:
1. Generate random numbers $U_1 \sim Uniform(0,1)$ and $U_2 \sim Uniform(0,1)$
2. Set:
$$X = \sqrt{-2\log U_1}\cos(2\pi U_2)$$
$$Y = \sqrt{-2\log U_1}\sin(2\pi U_2)$$

Box-Muller Algorithm v2: Suppose that (V_1, V_2) is uniformly distributed in the disk centered at $(0,0)$ with radius 1 and the random angle is $\theta \sim Uniform(0, 2\pi)$
1. Generate random numbers $U_1 \sim Uniform(0,1)$ and $U_2 \sim Uniform(0,1)$
2. Set $V_1 = 2U_1 - 1$, $V_2 = 2U_2 - 1$, $S = V_1^2 + V_2^2$ (V_1, V_2 are just X and Y coordinates sampled from Uniform(-1, 1)
3. If $S > 1$, return to Step 1 (S is the radius squared for a unit disk so it should be ≤ 1)
4. Return the independent unit normals
$$X = \sqrt{-\frac{2\log S}{S}}V_1$$
$$Y = \sqrt{-\frac{2\log S}{S}}V_2$$

Simple Sampling: Sample $X_1, X_2, \cdots, X_n$ independently from f, we can estimate the true parameter shown below by
$$\theta = E[\varphi(X)] = \int_S \varphi(X)f(x)\,dx$$
Simple Sampling Estimator:
$$\hat{\theta}_{SS} = \frac{1}{n}\sum_{i=1}^n \varphi(X_i)$$
Simple Sampling Exact Variance of $\hat{\theta}$ (Variance of Sample Mean):
$$Var(\hat{\theta}) = \frac{Var[\varphi(X)]}{n} = \frac{\int_S \varphi^2(X)f(x)dx - \theta^2}{n}$$
Simple Sampling Asymptotic Variance of $\hat{\theta}$:
$$\sigma^2 \equiv Var[\varphi(X)] = \int_S \varphi^2(x)f(x)\,dx - \theta^2$$
Simple Sampling Estimated Asymptotic Variance of $\hat{\theta}$: Note that this is not an unbiased estimate of σ^2
$$\hat{\sigma}_{SS}^2 = \frac{1}{n}\sum_{i=1}^n \varphi^2(X_i) - \hat{\theta_{SS}}^2$$
Simple Sampling Estimated Variance of $\hat{\theta}$ (Sample Variance):
$$\widehat{Var}(\hat{\theta}) = \frac{\hat{\sigma}_{SS}^2}{n}$$
Simple Sampling Asymptotic Confidence Interval for θ:
$$\left[\hat{\theta} - 1.96\frac{\hat{\sigma}_{SS}}{\sqrt{n}}, \hat{\theta} + 1.96\frac{\hat{\sigma}_{SS}}{\sqrt{n}}\right]$$

Importance Sampling: Sample $X_1, X_2, \cdots, X_n$ independently from g, we can estimate the true parameter shown below by
$$\theta = E_f[\varphi(X)] = \int_S \varphi(x)f(x)\,dx$$
$$\theta = \int_S \frac{\varphi(x)f(x)}{g(x)}g(x)\,dx = E_g\left[\frac{\varphi(X)f(X)}{g(X)}\right] = E_g[\varphi(Y)w(Y)]$$
Weighting Function:
$$w(y) = \frac{f(y)}{g(y)}$$
Importance Sampling Estimator: Unbiased estimator of θ
$$\hat{\theta}_{IS} = \frac{1}{n}\sum_{i=1}^n \frac{\varphi(x_i)f(x_i)}{g(x_i)} = \frac{1}{n}\sum_{i=1}^n \varphi(x_i)w(x_i)$$
Importance Sampling Exact Variance of $\hat{\theta}$ (Variance of Sample Mean):
$$Var(\hat{\theta}) = \frac{Var[\varphi(X)w(X)]}{n} = \frac{\int_S \frac{\varphi^2(x)f^2(x)}{g(x)}\,dx - \theta^2}{n}$$

Importance Sampling Asymptotic Variance of $\hat{\theta}$:
$$\sigma_{IS}^2 \equiv Var[\varphi(X)w(X)] = \int_S \frac{\varphi^2(x)f^2(x)}{g(x)}\,dx - \theta^2$$
Importance Sampling Estimated Asymptotic Variance of $\hat{\theta}$: Note that this is not an unbiased estimate of σ^2
$$\hat{\sigma}_{IS}^2 = \frac{1}{n}\sum_{i=1}^n \frac{\varphi^2(x)f^2(x)}{g(x)} - \hat{\theta_{IS}}^2$$
Importance Sampling Estimated Variance of $\hat{\theta}$ (Sample Variance):
$$\widehat{Var}(\hat{\theta}) = \frac{\hat{\sigma}_{IS}^2}{n}$$
Importance Sampling Asymptotic Confidence Interval for θ:
$$\left[\hat{\theta} - 1.96\frac{\hat{\sigma}_{IS}}{\sqrt{n}}, \hat{\theta} + 1.96\frac{\hat{\sigma}_{IS}}{\sqrt{n}}\right]$$
Optimal g: $g(x) \propto |\varphi(x)| \cdot f(x)$
- Asymptotic variance of $\hat{\theta}_{IS}$ with the proposal density g is exactly 0 if $\varphi(x) \ge 0$ for all $X \in S$

To find $g(x)$:
1. Let $h(x) = c|\varphi(x)|f(x)$
2. Let $1 = \int_S h(x) = \int_S c|\varphi(x)|f(x)$ and solve for c (Note that if we have the value for $I = \int_S |\varphi(x)|f(x)\,dx \Rightarrow c = \frac{1}{I}$
3. $g(x) = ch(x)$

Self-Normalizing Importance Sampling: We only know the distribution of f and g up to a normalising constant ($Z_f > 0, Z_g > 0$)
$$f(x) = \frac{\tilde{f}(x)}{Z_f}, \qquad g(x) = \frac{\tilde{g}(x)}{Z_g}$$
Generalised weights:
$$\tilde{w}(x) = \frac{\tilde{f}(x)}{\tilde{g}(x)}, \qquad for\ all\ x \in S$$
Self-normalised importance sampling estimator of $\theta = E_f[\varphi(X)] = \int_S \varphi(x)f(x)\,dx$:
$$\hat{\theta}_{SIS} = \frac{\sum_{i=1}^n \varphi(X_i)\tilde{w}(X_i)}{\sum_{i=1}^n \tilde{w}(X_i)}$$
Asymptotic variance of $\hat{\theta}_{SIS}$: Normally larger than the IS version because of the random denominator
$$\sigma_{SIS}^2 = E_g(w^2(X) \cdot [\varphi(X) - \theta]^2)$$
Where $w(x) = f(x)/g(x)$, it is the true weight

Exact Variance for $\hat{\theta}_{SIS}$: No closed form
- Note that $Var[\hat{\theta}_{SIS}] \neq \frac{\sigma_{SIS}^2}{n}$ (Not the same as simple sampling and importance sampling)

Estimator of the Variance of $\hat{\theta}_{SIS}$
$$\widehat{Var}(\hat{\theta}_{SIS}) = \frac{\hat{\sigma}_{SIS}^2}{n} = \frac{\sum_{i=1}^n \left\{\tilde{w}^2(X_i)\left[\varphi(X_i) - \hat{\theta}_{SIS}\right]^2\right\}}{\{\sum_{i=1}^n \tilde{w}(X_i)\}^2}$$
95% Asymptotic Confidence Interval
$$\left[\hat{\theta} - 1.96\sqrt{\frac{\hat{\sigma}_{SIS}^2}{n}}, \hat{\theta} + 1.96\sqrt{\frac{\hat{\sigma}_{SIS}^2}{n}}\right]$$

Calculus Results:
1. $\int_1^{+\infty} \frac{1}{x^p} = \frac{1}{-p+1}x^{-p+1}\Big|_1^{+\infty} = \begin{cases} < +\infty \text{ if p} > 1 \\ +\infty \ \ if\ p \le 1 \end{cases}$
2. $\int_0^1 \frac{1}{x^p} = \frac{1}{-p+1}x^{-p+1}\Big|_0^1 = \begin{cases} < +\infty \text{ if p} < 1 \\ +\infty \ \ if\ p \ge 1 \end{cases}$
3. $\int_0^1 \frac{e^x}{x^p} = \begin{cases} < +\infty \text{ if p} < 1 \\ +\infty \ \ if\ p \ge 1 \end{cases}$

Rare Event Estimation: When the p^* we want to estimate is small
Relative Standard Deviation = $\frac{asymptotic\ s.d}{p^*}$
- Checks the magnitude of the asymptotic sd of our estimator as compared to the actual value → If it is large, it means that the magnitude of the sd of our estimator is larger than the actual value and it will give a very bad estimate
- For a Bernoulli RV → $Relative\ s.d. = \frac{\sqrt{p(1-p)}}{p} = \sqrt{\frac{1-p}{p}}$ (Therefore, if the probability of it happening is low then the sd is high
- Consider using a density centered at the point where we need more points so that the probability is higher and lowering the relative sd
- Remember that we can take the φ as the indicator function to indicate $P(X_i > 4)$ for example.