ADM2303
Some Useful Formulas: Part 2

Prof.: Dr. Suren Phansalker

Some Discrete Distributions:

1. Geometric Distribution: (Enrichment)

RV, ‘X’ represents the # trials until the first success occurs.
p = probability of success
X~ Geom(p, x)
The distribution function of the RV, ‘X’ is:

$$P[X = x] = q^{(x-1)}\,p$$

$$E[X] = \mu = \frac{1}{1-q} = \frac{1}{p}$$

$$\{\sigma[X] = STD[X] = \sigma\} = \sqrt{\frac{q}{p^2}} = \frac{\sqrt{q}}{p}$$

2. X: RV distributed as a Hypergeometric Distribution: (Enrichment)

X ~ Hypergeom(N, n, p, x)
Obviously $N_1 + N_2 = N$. With:

$$p = \frac{N_1}{N} = \frac{M}{N} \text{ and } q = \frac{N_2}{N} = \frac{N-M}{M} \qquad \text{then,}$$

$$P[X = x] = \frac{\binom{M}{x}\binom{N-M}{n-x}}{\binom{N}{n}} = \frac{{}_{M}C_{x}\;{}_{(N-M)}C_{(n-x)}}{{}_{N}C_{n}} \qquad \text{and,}$$

$$\{E[X] = \mu\} = n\,p = \frac{n(M)}{N}$$

$$\{\sigma[X] = STD[X] = \sigma\} = \sqrt{npq}\sqrt{\left(\frac{N-n}{N-1}\right)}$$

3. **X: RV distributed as a Poisson Distribution:**
X ~ P[μ, x], where μ = Average occurrences over a given time horizon.

$P[X = x] = \dfrac{e^{-\mu}\mu^{x}}{x!}$

E[X] = μ

$\{STD[X] = \sigma[X] = \sigma\} = \sqrt{\mu}$

Continuous Distributions:

Basic Mathematical Properties and Relations: (Enrichment)

X: RV distributed continuously

Since continuous RV ‘X’ has infinite outcomes, and the Density Function, f(x) is defined, then:

$P[X = x] = 0$ ---1

$P[x \le X \le x + dx] = f(x)\,dx$ ---2

Since $0 \le f(x)\,dx \le 1,$ ---3

$f(x) \ge 0$, and ---4

$$\int_{-\infty}^{\infty} f(x)dx = 1 \qquad \text{---5}$$

$$\mu = E[X] = \int_{-\infty}^{\infty} xf(x)dx \qquad \text{---6}$$

$$\sigma^2[X] = Var[X] = \int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2 \qquad \text{---7}$$

$$\sigma[X] = STD[X] = \sqrt{\int_{-\infty}^{\infty} x^2 f(x)dx - \mu^2} = \sqrt{Var[X]} \qquad \text{---8}$$

Some Continuous Distribution/ Density Functions

1. Uniform Distribution/Density Function:

$f(x) = h$ and $a \le x \le b$

Here: $h = \dfrac{1}{(b-a)}$

and: $\mathbf{f}(x) = \dfrac{1}{(b-a)}$ and a ≤ x ≤ b

$$\mu = E[X] = \frac{(b+a)}{2}$$

Here $Var[X] = \dfrac{(b-a)^2}{12}$, and

$$\sigma[X] = STD[X] = \frac{(b-a)}{\sqrt{12}} = \frac{\sqrt{12}(b-a)}{12}$$

It can readily be seen that $P[x_1 \le x \le x_2] = (x_2 - x_1)\, h = \dfrac{(x_2 - x_1)}{(b-a)}$

2. Triangular Distribution/Density Functions:

See notes for Week9

3. Negative Exponential Distribution/Density Function:

$f(x) = \lambda e^{-\lambda x}$, and $0 \le x < \infty$ and $\lambda > 0$

with e = 2.71828…

$$\mu = E[X] = \frac{1}{\lambda}$$

$$\{\sigma^2[X] = Var[X]\} = \frac{1}{\lambda^2}$$

$$\{\sigma[X] = STD[X]\} = \frac{1}{\lambda}$$

$$\{P[X \ge x_0] = P[X > x_0]\} = e^{-\lambda x_0}$$

$P[x_1 \le (X = x) \le x_2] = [e^{-\lambda x_1} - e^{-\lambda x_2}]$

4. Normal/ Gaussian Distribution/Density Function:

$f(x) = \dfrac{1}{\sigma\sqrt{2\pi}} e^{\frac{-(x-\mu)^2}{2\sigma^2}}$, $-\infty < x < \infty$

$E(X) = \mu$, and

$Var(X) = \sigma^2$, and

$STD(X) = \sigma$

$X \sim N(\mu, \sigma)$ and with $Z = \dfrac{X - \mu}{\sigma}$, $E(Z) = 0, \sigma(Z) = 1$ {See Week10Pr file}

$A(z) = \int_0^z f(z)dz$, are the 'Table Values' from the Normal Distribution Table.

P[X > x] = $P\left[Z > \left(z = \frac{x-\mu}{\sigma}\right)\right]$

= 0.5 – A(z), if z > 0

{= 0.5 + A(z), if z < 0 }{Draw a diagram to understand this}

P[X < x] = $P\left[Z < \left(z = \frac{x-\mu}{\sigma}\right)\right]$

= 0.5 + A(z), if z > 0

{= 0.5 - A(z), if z < 0 }{Draw a diagram to understand this}

$$P[x_1 < X < x_2] = P\left[\left(z_1 = \frac{x_1-\mu}{\sigma}\right) < Z < \left(z_2 = \frac{x_2-\mu}{\sigma}\right)\right]$$

= $A(z_2) – A(z_1)$, $z_2 > z_1 > 0$

{Always draw a diagram to find the correct value.}

Some Famous Approximations:

1. Binomial to Normal Approximation: Used in this Course

You must use Continuity Correction.

X ~ b(n, p, x)
If
(n p) ≥ 5, and simultaneously
(n q) ≥ 5, then
$b(n, p, x) \to N\left(\mu = np, \sigma = \sqrt{npq}\right)$

2. Binomial to Poisson Approximation: (Enrichment)

No continuity correction is applied.

X ~ b(n, p, x)
If
p << 1 (very small)
n >> 1 (very large), but
(n p) = reasonable small with a finite value, then

$b(n, p, x) \to P\left(\mu = np, x\right)$

Central Limit Theorem (CLT)

1. Case 1 and Case 2 (See Notes) when 'σ' is known:

$$\bar{X} \sim N\left(E(\bar{X}) = \mu, \sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}\right)$$

Finite Population Correction Factor, FPCF, should always be used when sample size 'n' is relative large (n > N/100) with respect to the population size 'N'. If population size 'N', is not specified, it is always considered to be very large, and FPCF is considered to be 1.

$$FPCF = \sqrt{\frac{N-n}{N-1}}$$

When FPCF must be used,

$$\sigma(\bar{X}) = \frac{\sigma}{\sqrt{n}}\sqrt{\frac{N-n}{N-1}}$$

2. Case 3: when 'σ' is unknown but must use 's':

$$\bar{X} \sim t_{n-1}\left(E(\bar{X}) = \mu, s(\bar{X}) = \frac{s}{\sqrt{n}}\right)$$

3. Special Condition when 'σ' is unknown but must use 's':

If 'n' is large,

$$\bar{X} \sim t_{n-1}\left(E(\bar{X}) = \mu, s(\bar{X}) = \frac{s}{\sqrt{n}}\right) \to N\left(E(\bar{X}) = \mu, s(\bar{X}) = \frac{s}{\sqrt{n}}\right)$$

4. Brief Idea about 't' Distribution: (Enrichmet)

If $\bar{x}_0$ is a specific value of the RV, sample mean, or $\bar{X}$, and sample size is 'n', then with $s(\bar{X}) = \frac{s}{\sqrt{n}}$

$$P[\bar{X} > \bar{x}_0] = P\left[t_{n-1} > \left(t_0 = \frac{\bar{x}_0 - \mu}{s(\bar{X})}\right)\right]$$

$$P[\overline{X} > \overline{x}_0] = P\left[t_{n-1} > \left(t_0 = \frac{\overline{x}_0 - \mu}{\left(\frac{s}{\sqrt{n}}\right)} \right) \right] = A(t_0), t_0 > 0$$

This A(t_0), is the 'Right Tail' area in the 't' table.
{Always draw a diagram for clear understanding of the problem.}

5. CLT for Sample Proportion, $\hat{p}$: Used in the Course

Here the theory is mainly based on 'Binomial' to 'Normal' approximation:

p: Population Proportion,
n: Sample Size

$\hat{p} = \frac{X}{n}$, where X ~ b(n, p, x), then if simultaneously,
(n p) > 5, and (n q) > 5

$$\hat{p} \sim N\left(E(\hat{p}) = p, \sigma(\hat{p}) = \left(\sqrt{\frac{pq}{n}} \right) \right)$$

Finite Population Correction Factor, FPCF, should always be used when sample size, 'n', is relative large (n > N/100) with respect to the population size 'N'. If population size 'N', is not specified, it is always considered to be very large, and FPCF is considered to be 1.

$$FPCF = \sqrt{\frac{N-n}{N-1}}$$

When FPCF must be used,

$$\sigma(\hat{p}) = \sqrt{\frac{pq}{n}} \sqrt{\frac{N-n}{N-1}}$$

Statistical Quality Control:

$$\overline{\overline{X}} = \frac{\sum_{i=1}^{i=k} \overline{X}_i}{k} = E[\overline{X}] = \mu \qquad \text{, and } \overline{R} = \frac{\sum_{i=1}^{i=k} R_i}{k}$$

The best estimate of $\sigma(\overline{X})$ is given by: $\frac{A_2 \overline{R}}{3}$ where A_2 is famous Shewhart factor, developed by Walter Shewhart and is usually given in an appropriate table. Then the Lower Control Limit (LCL) and Upper Control Limit (UCL) for the X-Bar chart are given respectively by:

$$LCL = \overline{\overline{X}} - 3\frac{A_2}{3}\overline{R} = \overline{\overline{X}} - A_2\overline{R} \qquad \text{and}$$

$$UCL = \overline{\overline{X}} + 3\frac{A_2}{3}\overline{R} = \overline{\overline{X}} + A_2\overline{R}$$

The necessary LCL and UCL for the R-chart are given as follows:

$LCL = D_3\overline{R}$, and

$UCL = D_4\overline{R}$

These Control Chart Factors, ‘A_2’, ‘D_3’ and ‘D_4’ are given in the Shewhart Table.