Discrete distributions worksheet

Name of special distribution (X)	Properties	What X represents and its range (R_X)	Parameters	Probability $P(X = x)$	Expected value $E(X)$	Variance $Var(X)$
Binomial distribution	○ You perform $\boldsymbol{n}$ identical experiments and you know $\boldsymbol{n}$ before starting ○ Each experiment or trial independent of each other ○ The probability of success, $\boldsymbol{p}$ (or seeing a particular event) is same for every trial	The number of times success is observed $\boldsymbol{R_X}$: {0,1,2…,n}	n, p	$\binom{n}{x} p^x (1-p)^{n-x}$	np	$np(1-p)$
Hypergeometric distribution	○ I sample k objects from a finite population with b success and r failures ○ Each sample of k objects is equally likely to be chosen	The number successes observed $\boldsymbol{R_X}$: $\{0,1,2,..min\{b,k\}\}$	k, b, r	$\frac{\binom{b}{x}\binom{r}{k-x}}{\binom{b+r}{k}}$	$\frac{kb}{b+r}$	$k * \frac{b}{b+r} * \frac{r}{b+r} * \frac{b+r-k}{b+r-1}$
Negative binomial distribution	○ You perform identical experiments and you do not know $\boldsymbol{n}$ before starting, rather you keep performing experiments until you observe m-th success ○ Each experiment or trial independent of each other ○ The probability of success, p is same for every trial	The number of trials before m −th success is observed $\boldsymbol{R_X}$: $\{m, m+1, \ldots \infty\}$	m, p	$\begin{cases} \binom{x-1}{m-1} p^m (1-p)^{x-m} \\ 0 \end{cases}$	$\frac{m}{p}$	$\frac{m(1-p)}{p^2}$
Poisson distribution	○ An event occurs at an average rate of λ such that: ○ Occurrences of the event are independent of each other ○ More than one of these events cannot occur simultaneously	The number of events $\boldsymbol{R_X}$: $\{0,1,2 \ldots \infty\}$	λ	$\frac{e^{-\lambda}\lambda^x}{x!}$	λ	λ

Continuous distributions worksheet

Name of special distribution (X)	Properties	Range (R_X) and Parameters	PDF (it is not probability X=x) $f(x)$	CDF $F(x) = P(X \le x)$	Expected value $E(X)$	Variance $Var(X)$
Uniform distribution	○ If all possible values of X are equally likely to occur ○ And, all values fall in the range $[a, b]$	R_X: $[a, b]$ Parameters: a, b	$f(x) = \begin{cases} \frac{1}{b-a} & a \le x \le b \\ 0 & otherwise \end{cases}$	$f(x) = \begin{cases} 0 & x \le a \\ \frac{x-a}{b-a} & a \le x \le b \\ 1 & x \ge b \end{cases}$	$(b+a)/2$	$\frac{(b-a)^2}{12}$
Exponential distribution	○ X represents the waiting time between Poisson distribution events ○ X has property of being memoryless which means $P(X \ge x + x_0 \| X \ge x_0) = P(X \ge x)$	R_X: $[0, \infty)$ Parameters: λ	$f(x) = \begin{cases} \lambda e^{-\lambda x} & x \ge 0 \\ 0 & otherwise \end{cases}$	$F(x) = 1 = e^{\{-\lambda x\}}$	$1/\lambda$	$\frac{1}{\lambda^2}$
Normal distribution	○ When the pdf of X has a bell curve shape	R_X: $(-\infty, \infty)$ Parameters: μ, σ	$f(x) = \frac{1}{\sqrt{\{2\pi\}}\sigma} e^{-(x-\mu)^2/2\sigma^2}$	Use z-table or R	μ	σ^2
Gamma distribution	○ General case of Exponential distribution	R_X: $[0, \infty)$ Parameters: α, λ	$f(x) = \begin{cases} \frac{1}{\Gamma(\alpha)} \lambda^\alpha x^{\alpha-1} e^{-\lambda x} & x \ge 0 \\ 0 & otherwise \end{cases}$	Use R	α/λ	$\frac{\alpha}{\lambda^2}$

We will learn about two more continuous distributions later in the course: t-distribution and chi-squared distribution

$\Gamma(.)$ is the Gamma function with properties: (a) $\Gamma(\alpha) = (\alpha - 1)\Gamma(\alpha - 1)$ (b) $\Gamma(n) = (n-1)!$ if n is positive integer (c) $\Gamma(1/2) = \sqrt{\pi}$

Given: $X_1, X_2, \ldots, X_n$ are random samples i.e. they are **independent and identically distributed**

A confidence interval (CI) gives a plausible range for a population parameter p by using outcomes from a sample

A prediction interval (PI) gives a plausible range for a single future prediction value

A tolerance interval (TI) gives a plausible range which contains at least k% of the entire population

Assumption	X_i's are from **normal distribution** with unknown mean and unknown variance	X_i's are from **normal distribution** with unknown mean and unknown variance	X_i's are from **normal distribution** with unknown mean and unknown variance and n<40
Interval	Confidence interval on population variance	Prediction interval for X of a single individual	Tolerance interval containing at least $k\%$ of population
Notations	n: $\chi^2_{\alpha,n-1}$: s:	$\bar{x}$: n: s: $t_{\frac{\alpha}{2},n-1}$:	$\bar{x}$: n: s: $C_{\alpha,k}$:
$(1-\alpha)100\%$ two-sided interval	Lower limit: $(n-1)s^2/\chi^2_{\frac{\alpha}{2},n-1}$ Upper limit: $(n-1)s^2/\chi^2_{1-\frac{\alpha}{2},n-1}$	Lower limit: $\bar{x} - t_{\frac{\alpha}{2},n-1} * s * \sqrt{1+\frac{1}{n}}$ Upper limit: $\bar{x} + t_{\frac{\alpha}{2},n-1} * s * \sqrt{1+\frac{1}{n}}$	Lower limit: $\bar{x} - C_{\alpha,k}s$ Upper limit: $\bar{x} + C_{\alpha,k}s$ (two sided C values)
$(1-\alpha)100\%$ one-sided upper bound interval	Lower limit: $-\infty$ Upper limit: $(n-1)s^2/\chi^2_{1-\alpha,n-1}$	Lower limit: $-\infty$ Upper limit: $\bar{x} + t_{\alpha,n-1} * s * \sqrt{1+\frac{1}{n}}$	Lower limit: $-\infty$ Upper limit: $\bar{x} + C_{\alpha,k}s$ One sided C value
$(1-\alpha)100\%$ one-sided lower bound interval	Lower limit: $(n-1)s^2/\chi^2_{\alpha,n-1}$ Upper limit: $+\infty$	Lower limit: $\bar{x} - t_{\alpha,n-1} * s * \sqrt{1+\frac{1}{n}}$ Upper limit: $+\infty$	Lower limit: $\bar{x} - C_{\alpha,k}s$ Upper limit: $+\infty$ One sided C value

Population Parameter	Population mean
Null hypothesis	$\mu = \mu_0$
Test statistic	$\frac{x-\mu_0}{s/\sqrt{n}}$ OR $\frac{x-\mu_0}{\sigma/\sqrt{n}}$ based on if population variance is known
Case 1	$n > 40$ OR $n < 40$ AND population distribution is normal AND population variance is known
Case 2	$n < 40$ AND population distribution is normal AND variance not known
Otherwise	Consult a knowledgeable statistician

Alternate Hypothesis test type	**Rejection region Case 1**	**Rejection region Case 2**
$H_a\colon \mu > \mu_0$ Or $H_a\colon p > p_0$	(z_α, ∞)	$(t_{\alpha,n-1}, \infty)$
$H_a\colon \mu < \mu_0$ Or $H_a\colon p < p_0$	$(-\infty, -z_\alpha)$	$(-\infty, -t_{\alpha,n-1})$
$H_a\colon \mu \neq \mu_0$ Or $H_a\colon p \neq p_0$	$(-\infty, -z_{\frac{\alpha}{2}}) \cup (z_{\frac{\alpha}{2}}, \infty)$	$(-\infty, -t_{\frac{\alpha}{2},n-1}) \cup (t_{\frac{\alpha}{2},n-1}, \infty)$

Reject if test statistic belongs to rejection region

Population Parameter	Sample proportion
Null hypothesis	$p = p_0$
Test statistic	$\frac{p' - p_0}{\sqrt{p_0(1-p_0)/n}}$
Case 1	$np_0 > 10$ AND $n(1-p_0) > 10$
Otherwise	Consult a knowledgeable statistician

Hypothesis test type	**P-values Case 1**
$H_a\colon \mu > \mu_0$ Or $H_a\colon p > p_0$	$1 - \phi(z)$
$H_a\colon \mu < \mu_0$ Or $H_a\colon p < p_0$	$\phi(z)$
$H_a\colon \mu \neq \mu_0$ Or $H_a\colon p \neq p_0$	$2[1 - \phi(\lvert z \rvert)]$

Reject if $\alpha > P$ value