Discrete Probability Distributions

A CEE3030 lecture prepared by

Gilberto E. Urroz

February 2006

Reference

The subjects presented are taken from the Maple worksheed entitled

DiscreteProbabilityDistributions

available for download in the class schedule

Quick review of concepts for discrete random variables - 1

- Let *X* be a discrete random variable, then
  - *f(x) = P(X=x)* is the *probability mass function (pmf)*
  - *F(x) = P(X≤x)* = $\sum_{u \le x} f(u)$ = *cumulative distribution function (CDF)*
- Calculation of probabilities
  - *P(X < x) = F(x-1)* -- *P(X≤x) = F(x)*
  - *P(X > x) = 1-F(x)* -- *P(X≥x) =1-F(x-1)*
  - *P(a < X < b) = F(b-1)-F(a)*
  - *P(a ≤X< b) = F(b-1)-F(a-1)*
  - *P(a < X ≤b) = F(b)-F(a)*
  - *P(a ≤X ≤b) = F(b)-F(a-1)*

Quick review of concepts for discrete random variables - 2

- Let *X* be a discrete random variable, then
  - *f(x) = P(X=x)* is the *probability mass function (pmf)*
- Calculation of measures
  - Mean, $\mu = \sum_{i=1}^{n} x_i \cdot f(x_i)$
  - Variance, $\sigma^2 = \sum_{i=1}^{n} (x_i - \mu)^2 \cdot f(x_i)$
  - Skewness $\alpha_3 = \frac{1}{\sigma^3} \sum_{i=1}^{n} (x_i - \mu)^3 \cdot f(x_i)$
  - Kurtosis $\alpha_4 = \frac{1}{\sigma^4} \sum_{i=1}^{n} (x_i - \mu)^4 \cdot f(x_i)$

Discrete distributions in Maple

- Use the command: *?Statistics,Distributions* for a list of available distributions
- Discrete distributions of interest are:

•Bernoulli	Bernoulli distribution
•Binomial	binomial distribution
•DiscreteUniform	discrete uniform distribution
•EmpiricalDistribution	empirical distribution
•Geometric	geometric distribution
•Hypergeometric	hypergeometric distribution
•NegativeBinomial	negative binomial (Pascal) dist.
•Poisson	Poisson distribution
•ProbabilityTable	probability table

Using Maple Statistics *package to define a discrete random variable*

- To load the *Statistics* package use: *with(Statistics)*
- Use *?<distribution name>* for help
  - e.g., *?Geometric*
- Define a random variable with distribution name and appropriate parameters with function *RandomVariable*
  - e.g., *X := RandomVariable(Binomial(n,p))*
  - e.g., *X := RandomVariable(Poison(3.2))*

Calculating measures of a distribution - 1

- After defining a random variable *X* in Maple, you can calculate the following measures:

- μ := *Mean(X)*
- $\sigma 2$:= *Variance(X)*
- σ := *StandardDeviation(X)*
- $\alpha 3$:= *Skewness(X)*
- $\alpha 4$:= *Kurtosis(X)*

Calculating measures of a distribution - 2

- To obtain floating-point (decimal) results for the measures of a distribution you may use:

- μ := *evalf(Mean(X))*
- $\sigma 2$:= *evalf(Variance(X))*
- σ := *evalf(StandardDeviation(X))*
- $\alpha 3$:= *evalf(Skewness(X))*
- $\alpha 4$:= *evalf(Kurtosis(X))*

Calculating probabilities - 1

- To calculate probabilities use the following basic functions:
  - *ProbabilityFunction(X,a)* for the *pmf*, i.e., $f(a)=P(X=a)$
  - *CDF(X,a)* for the *CDF*, i.e., $F(a) = P(X \le a)$

Calculating probabilities - 2

- To calculate more complex probabilities use function *CDF* as follows:
  - $P(X < x) = F(x-1)$ => use *CDF(X,x-1)*
  - $P(X > x) = 1-F(x)$ => use *1-CDF(X,x-1)*
  - $P(X \ge x) = 1-F(x-1)$ => use *1-CDF(X,x-1)*
  - $P(a < X < b) = F(b-1)-F(a)$ => use *CDF(X,b-1)-CDF(x,a)*
  - $P(a \le X < b) = F(b-1)-F(a-1)$ => use *CDF(X,b-1)-CDF(x,a-1)*
  - $P(a < X \le b) = F(b)-F(a)$ => use *CDF(X,b)-CDF(x,a)*
  - $P(a \le X \le b) = F(b)-F(a-1)$ => use *CDF(X,b)-CDF(x,a-1)*
-

The Bernoulli distribution

- Random variable *X* can take only the values $x = 0$ and $x = 1$
- Probability mass function: with $0 < p < 1$

$$f(x) = \begin{cases} 1-p & x=0 \\ p & x=1 \\ 0 & otherwise \end{cases}$$

- Possible association of the values of *x*:

Variable *X*	*X=0* equivalent	*X=1* equivalent
Binary logical	No	Yes
Voltage level	Low voltage	High voltage
Sucess/failure	Failure	Success

Measures of the Bernoulli distribution

- $$\mu = p$$
- $$\sigma^2 = p \cdot (1-p)$$
- $$\sigma = \sqrt{p(1-p)}$$
- $$\alpha_3 = \frac{1-2p}{\sqrt{p(1-p)}}$$
- $$\alpha_4 = \frac{1-3p+3p^2}{p(1-p)}$$

The binomial distribution: X~B(n,p)

- Consider n repetitions of a Bernoulli process with parameter p
- Let X = number of "successes" in n repetitions
- Probability mass function

$$f(x)=\binom{n}{x}p^x(1-p)^{n-x},\ for\ x=0,1,\ldots,n$$

- Binomial coefficient: $\binom{n}{x}=\frac{n!}{x!(n-x)!}$

Measures of the binomial distribution

- $\mu=np$
- $\sigma^2=np(1-p)$
- $\sigma=\sqrt{np(1-p)}$
- $\alpha_3=\frac{1-2p}{\sqrt{np(1-p)}}$
- $\alpha_4=long\ expression\ ,\ see\ worksheet$

Approximating the binomial distribution with the normal distribution, X~N(μ,σ)

- Applies for relatively large values of n and relatively small values of p so that

  $np \geq 5$ or $n(1\text{-}p) \geq 5$

- Use *X := RandomVariable(Normal(μ,σ))* to define a normal random variable (continuous)
- Main reason for the approximation: to avoid calculating large factorial values – No longer an impediment with modern calculators and software

The Poisson distribution

- Used to define discrete random variable X = number of occurrences of a certain phenomena per unit time, unit length, etc.
- Probability mass function

$$f(x)=\frac{e^{-\lambda}\lambda^x}{x!},\ for\ x=0,1,\ldots$$

- Parameter λ represents the average number of occurrence per unit time, length, etc.

Measures of the Poisson distribution

- $\mu=\lambda$
- $\sigma^2=\lambda$
- $\sigma=\sqrt{\lambda}$
- $\alpha_3=\frac{1}{\sqrt{\lambda}}$
- $\alpha_4=\frac{3\lambda+1}{\lambda}$

Poisson distribution with scaling

- Let X = number of occurrences of a phenomenon, say, per unit time
- Let σ = average number of occurrences per unit time
- Let T = period of interest for the analysis
- Use $\lambda=\sigma T$ as the parameter in the Poisson distribution
- See example of scaling in worksheet