# B671-672 Supplemental Notes 2
# Hypergeometric, Binomial, Poisson and Multinomial Random Variables and Borel Sets

## 1 Binomial Approximation to the Hypergeometric

Recall that the Hypergeometric Distribution is

$$f(x) = \begin{cases} \frac{\binom{n}{x}(D)_x(N-D)_{n-x}}{(N)_n} = \frac{\binom{D}{x}\binom{N-D}{n-x}}{\binom{N}{n}} & x = \max(0, n-N+D), \ldots, \min(n, D) \\ 0 & \text{elsewhere} \end{cases}$$

which we may write as

$$f(x) = \frac{\prod_{i=0}^{x-1}(D-i)\prod_{i=0}^{n-x-1}(N-D-i)}{\prod_{i=0}^{n-1}(N-i)}$$

Since there are $n$ terms in the numerator and the denominator we may divide both by $N^n$ to obtain

$$f(x) = \frac{\prod_{i=0}^{x-1}\left(\frac{D}{N}-\frac{i}{N}\right)\prod_{i=0}^{n-x-1}\left(1-\frac{D}{N}-\frac{i}{N}\right)}{\prod_{i=0}^{n-1}\left(1-\frac{i}{N}\right)}$$

Since

$$\left(\frac{D}{N}-\frac{x-1}{N}\right) \leq \left(\frac{D}{N}-\frac{i}{N}\right) \leq \left(\frac{D}{N}\right) \text{ for } i = 0, 1, \ldots, x-1$$

$$\left(1-\frac{D}{N}-\frac{n-x-1}{N}\right) \leq \left(1-\frac{D}{N}-\frac{i}{N}\right) \leq \left(1-\frac{D}{N}\right) \text{ for } i = 0, 1, \ldots, n-x-1$$

$$\left(1-\frac{n-1}{N}\right) \leq \left(1-\frac{i}{N}\right) \leq 1 \text{ for } i = 0, 1, \ldots, n-1$$

we have that

$$\left(\frac{D}{N}-\frac{x-1}{N}\right)^x \le \prod_{i=0}^{x-1}\left(\frac{D}{N}-\frac{i}{N}\right) \le \left(\frac{D}{N}\right)^x$$

$$\left(1-\frac{D}{N}-\frac{n-x-1}{N}\right)^{n-x} \le \prod_{i=0}^{n-x-1}\left(1-\frac{D}{N}-\frac{i}{N}\right) \le \left(1-\frac{D}{N}\right)^{n-x}$$

$$\left(1-\frac{n-1}{N}\right)^n \le \prod_{i=0}^{n-1}\left(1-\frac{i}{N}\right) \le 1$$

and it follows that

$$\binom{n}{x}\left(\frac{D}{N}-\frac{x-1}{N}\right)^x\left(1-\frac{D}{N}-\frac{n-x-1}{N}\right)^{n-x} \le f(x) \le \frac{\binom{n}{x}\left(\frac{D}{N}\right)^x\left(1-\frac{D}{N}\right)^{n-x}}{\left(1-\frac{n-1}{N}\right)^n}$$

If we assume that $\lim_{N\to\infty}\left(\frac{D}{N}\right)=p$ and that $n$ and $x$ are fixed then

$$\lim_{N\to\infty}\binom{n}{x}\left(\frac{D}{N}-\frac{x-1}{N}\right)^x\left(1-\frac{D}{N}-\frac{n-x-1}{N}\right)^{n-x} = \binom{n}{x}p^x(1-p)^{n-x}$$

and

$$\lim_{N\to\infty}\frac{\binom{n}{x}\left(\frac{D}{N}\right)^x\left(1-\frac{D}{N}\right)^{n-x}}{\left(1-\frac{n-1}{N}\right)^n} = \binom{n}{x}p^x(1-p)^{n-x}$$

It follows that

$$\lim_{N\to\infty}\frac{\binom{n}{x}(D)_x(N-D)_{n-x}}{(N)_n} = \binom{n}{x}p^x(1-p)^{n-x}$$

so that the hypergeometric distribution can be approximated by the binomial with $p=\frac{D}{N}$

## 2 Poisson Approximation to the Binomial

The probability of $x$ successes in $n$ Bernoulli trials with $n$ trials and probability of success $p_n$ on each trial is given by the binomial distribution i.e.

$$P_n(X = x) = \binom{n}{x} p_n^x (1 - p_n)^{n-x} \text{ for } x = 0, 1, 2, \ldots, n$$

Suppose now that $n \uparrow$ such that

$$\lim_{n\to\infty} np_n = \lim_{n\to\infty} \lambda_n = \lambda > 0$$

The binomial distribution can then be written as

$$\begin{aligned} P_n(X = x) &= \binom{n}{x} p_n^x (1 - p_n)^{n-x} \\ &= \frac{n(n-1)\cdots(n-x+1)}{x!} \left(\frac{\lambda_n}{n}\right)^x \left(1 - \frac{\lambda_n}{n}\right)^{n-x} \\ &= \prod_{i=0}^{x-1} \left(1 - \frac{i}{n}\right) \frac{\lambda_n^x}{x!} \left(1 - \frac{\lambda_n}{n}\right)^{n-x} \end{aligned}$$

Hence

$$\left(1 - \frac{x-1}{n}\right)^x \frac{\lambda_n^x}{x!} \left(1 - \frac{\lambda_n}{n}\right)^n \le P_n(x) \le \frac{\lambda_n^x}{x!} \left(1 - \frac{\lambda_n}{n}\right)^n \left(1 - \frac{\lambda_n}{n}\right)^{-x}$$

Now for fixed $x$

$$L_n(x) = \left(1 - \frac{x-1}{n}\right)^x \frac{\lambda_n^x}{x!} \left(1 - \frac{\lambda_n}{n}\right)^n \longrightarrow \frac{\lambda^x}{x!} e^{-\lambda}$$

and

$$U_n(x) = \frac{\lambda_n^x}{x!} \left(1 - \frac{\lambda_n}{n}\right)^n \left(1 - \frac{\lambda_n}{n}\right)^{-x} \longrightarrow \frac{\lambda^x}{x!} e^{-\lambda}$$

so that

$$\lim_{n\to\infty} L_n(x) = \lim_{n\to\infty} U_n(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

Hence

$$\lim_{n\to\infty} P_n(x) = \frac{\lambda^x}{x!} e^{-\lambda}$$

i.e. the binomial distribution approaches the Poisson as $np_n \to \lambda$ and hence can be used to approximate the binomial under this condition.

## 3 Multivariate Hypergeometric and Multinomial Distributions

Consider a population of $N$ individuals each classified into one of $k$ mutually exclusive categories $C_1, C_2, \ldots, C_k$. Suppose that there are $D_i$ individuals in category $C_i$ for $i = 1, 2, \ldots, k$. Note that $\sum_{i=1}^{k} D_i = N$. If we draw a random sample of size $n$ without replacement from this population the probability of $x_i$ in the sample from category $C_i$, $i = 1, 2, \ldots, k$ is

$$P(x_1, x_2, \ldots, x_k) = \frac{n!}{x_1! x_2! \cdots x_k!} \frac{(D_1)_{x_1}(D_2)_{x_2} \cdots (D_k)_{x_k}}{(N)_n}$$

which is called the multivariate hypergeometric distribution with parameters $D_1, D_2, \ldots, D_k$. It is not widely used since the multinomial distribution provides an excellent approximation. Rewrite the distribution as

$$P(x_1, x_2, \ldots, x_k) = \frac{n!}{\prod_{i=1}^{k} x_i!} \frac{\prod_{i=1}^{k} \prod_{j=0}^{x_i-1} (D_i - j)}{\prod_{l=0}^{n-1} (N - l)}$$

Dividing the numerator and denominator by $N^n$ as in the development of the binomial approximation to the hypergeometric yields

$$P(x_1, x_2, \ldots, x_k) = \frac{n!}{\prod_{i=1}^{k} x_i!} \frac{\prod_{i=1}^{k} \prod_{j=0}^{x_i-1} \left(\frac{D_i}{N} - \frac{j}{N}\right)}{\prod_{l=0}^{n-1} \left(1 - \frac{l}{N}\right)}$$

Since

$$\left(\frac{D_i}{N} - \frac{x_i - 1}{N}\right) \leq \left(\frac{D_i}{N} - \frac{j}{N}\right) \leq \left(\frac{D_i}{N}\right) \quad \text{for} \quad j = 0, 1, \ldots, x_i - 1 \quad i = 1, 2, \ldots, k$$

$$\left(1 - \frac{n-1}{N}\right) \leq \left(1 - \frac{l}{N}\right) \leq 1 \quad \text{for} \quad l = 0, 1, \ldots, n - 1$$

we have that

$$\prod_{i=1}^{k}\left(\frac{D_i}{N}-\frac{x_i-1}{N}\right)^{x_i} \le \prod_{i=1}^{k}\prod_{j=0}^{x_i-1}\left(\frac{D_i}{N}-\frac{j}{N}\right) \le \prod_{i=1}^{k}\left(\frac{D_i}{N}\right)^{x_i}$$

and

$$\left(1-\frac{n-1}{N}\right)^{n} \le \prod_{i=0}^{n-1}\left(1-\frac{l}{N}\right) \le 1$$

Thus the multivariate hypergeometric distribution is bounded below by

$$\frac{n!}{x_1!x_2!\cdots x_k!}\prod_{i=1}^{k}\left(\frac{D_i}{N}-\frac{x_i-1}{N}\right)^{x_i}$$

and is bounded above by

$$\frac{n!}{x_1!x_2!\cdots x_k!}\frac{\prod_{i=1}^{k}\left(\frac{D_i}{N}\right)^{x_i}}{\left(1-\frac{n-1}{N}\right)^{n}}$$

Thus if $n$ is fixed and $\lim_{N\to\infty}\frac{D_i}{N}=p_i$ for each $i$ then

$$\lim_{N\to\infty} P(x_1,x_2,\ldots,x_k)=\frac{n!}{x_1!x_2!\cdots x_k!}\prod_{i=1}^{k}p_i^{x_i}$$

It follows that the multivariate hypergeometric distribution can be approximated by the multinomial distribution with $p_i=\frac{D_i}{N}$ for $i=1,2,\ldots,k$.

# 4 Borel Sets and Measurable Functions

## 4.1 Necessity for Borel Sets

Let $\Omega = [0, 1)$ and let $P(E)$ be the length of $E$ i.e. the uniform probability measure. It is not possible for $P$ to be defined for all subsets of $\Omega$ in such a way that $P$ satisfies $0 \le P(E) \le 1$, $P(\Omega) = 1$, $P$ is countably additive, and $P(E)$ is equal to the "length" of $E$.

To show this a non-measurable set is constructed by the following procedure:

(1) Define an equivalence relation on $\Omega = [0, 1)$ such that

$$\Omega = [0, 1) = \cup_{t \in T} E_t$$

where the $E_t$ are disjoint. The equivalence relation is defined by

$$x \sim y \iff \text{mod}_1(x + r) = y \;\text{ where } r \text{ is a rational number}$$

where

$$\text{mod}_1(x_1 + x_2) = \begin{cases} x_1 + x_2 & \text{if } x_1 + x_2 \le 1 \\ x_1 + x_2 - 1 & \text{if } x_1 + x_2 > 1 \end{cases} \quad \text{for } x_1, x_2 \in [0, 1)$$

(2) Use the Axiom of Choice to select a set $F$ containing exactly one point from each of the $E_t$.

(3) Order the rational numbers as

$$r_0, r_1, r_2, \ldots \;\text{ where } r_0 = 0$$

and define the sets $F_i$ by

$$F_i = \text{mod}_1(F + r_i)$$

You can show that the $F_i$ are mutually exclusive and that

$$\cup_{i=0}^{\infty} F_i = \Omega = [0, 1)$$

(4) If $F$ is measurable so are each of the $F_i$ and

$$P(F) = P(F_i)$$

It then follows that

$$\begin{aligned} 1 &= P(\Omega) \\ &= P([0, 1)) \\ &= P(\cup_{i=0}^{\infty} F_i) \\ &= \sum_{i=0}^{\infty} P(F_i) \\ &= \sum_{i=0}^{\infty} P(F) \end{aligned}$$

Thus $F$ is not measureable i.e. no value $P(F)$ can be assigned to $F$.

**Reference:** Counterexamples in Probability and Statistics, (1986) Romano and Siegel, Wadsworth and Brooks/Cole

## 4.2 Measurable Functions and Random Variables

In the study of probablity and its applications to statistics we need to have a collection of random variables (measurable functions) large enough to ensure that probabilities are well defined.

Recall that most of classical analysis (calculus, etc.) deals with continuous functions and limits of sequences of continuous functions. Since limits of sequences of continuous functions are not necessarily continuous and also since a sequence of continuous functions may not tend to a finite limit it is convenient to extend slightly the definition of the functions under consideration by allowing functions to take the values $\pm\infty$ and to extend the notion of a Borel $\sigma$-field to be the smallest $\sigma$-field generated by the sets $\{-\infty\}$, $\{+\infty\}$ and $\mathcal{B}$ where $\mathcal{B}$ is the class of all Borel sets. With this extension the collection of measurable functions from

$$(\Omega, \mathcal{W}) \text{ to } (\overline{R}, \overline{\mathcal{B}})$$

is closed under the following "usual" operations:

- arithmetic operations (sums, products)
- infima and suprema
- pointwise limits of sequences of functions

There are now two definitions of measurable functions:

**Constructive:** A measurable function is the limit of a convergent sequence of simple functions where a simple function is a function of the form

$$X_n = \sum_{i=1}^{n} x_j I_{E_j}$$

where the $E_j$ are measurable sets (i.e. $E_j \in \mathcal{W}$).

**Descriptive:** A measurable function is a function such that the inverse image of any Borel set in $\mathcal{B}$ is a measurable set in $\mathcal{W}$.

**Theorem 4.1** The constructive and descriptive definitions are equivalent. Moreover the set of all measurable functions is closed under the "usual" operations of analysis

**Proof:** See Loeve page 107-108.

**Theorem 4.2** A continuous function of a measurable function is measurable. The class of functions containing all continuous functions and closed under limits are called **Baire** functions. It follows that Baire functions of measurable functions are also measurable.

**Proof:** See Loeve page 108-109.

Reference: Loeve, M. (1960) Probability Theory (Second Edition) Van Nostrand

**Theorem 4.3** A function $X = (X_1, X_2, \ldots, X_n$ from $\Omega$ to $\overline{R}$ is measurable if and only if each of the coordinates $X_1, X_2, \ldots, X_n$ is measurable.

**Proof:** See Loeve page 108-109.

Reference: Loeve, M. (1960) Probability Theory (Second Edition) Van Nostrand

It follows that the class of measurable functions or random variables is rich enough to include all random variables of potential interest. (Much like the fact that the class of Borel sets is rich enough to contain all events of interest and still be compatible with the need for probabilities to be countably additive).

**example 4.1** Define the signum function by

$$\operatorname{sgm}(x) = \begin{cases} +1 & \text{if } x > 0 \\ 0 & \text{if } x + 0 \\ -1 & \text{if } x < 0 \end{cases}$$

and note that the signum function is not continuous at $x = 0$.

Define the function $f_n(x)$ by

$$f_n(x) = \begin{cases} \min(1, nx) & x \geq 0 \\ \max(-1, nx) & x < 0 \end{cases}$$

Note that $f_n(0) = 0$ and that $f_n(x)$ is continuous for every $x$. However

$$\lim_{n\to\infty} f_n(x) = \operatorname{sgm}(x)$$

which is not continuous at $x = 0$ i.e. the limit of a sequence of continuous functions is not necessarily continuous.

**Reference:** Counterexamples in Analysis,(1964) Gelbaum and Olsted, Holden-Day Inc. page 77

**example 4.2** Define the function $f_n(x)$ by

$$f_n(x) = \begin{cases} \min\left(n \, , \, \frac{1}{x}\right) & \text{if } 0 < x \leq 1 \\ 0 & \text{if } x = 0 \end{cases}$$

Then $f_n(x)$ is bounded on the interval $[0, 1]$. However the function

$$f(x) = \lim_{n\to\infty} f_n(x) = \begin{cases} \frac{1}{x} & \text{if } 0 < x \leq 1 \\ 0 & \text{if } x = 0 \end{cases}$$

is unbounded. Thus the limit of a sequence of bounded functions is not necessarily bounded.

**Reference:** Counterexamples in Analysis,(1964) Gelbaum and Olsted, Holden-Day Inc. page 77