Week 8

Interchanging Sum and Differential

Suppose $\sum_{x=0}^{\infty} h(x,\theta)$ exists (i.e. converges pointwise) $\forall \ \theta \in (a,b)$

Moreover, assume

    i) $\frac{d}{d\theta} h(x,\theta)$ is continuous in θ for each x

    2) $\sum_{x=0}^{\infty} \frac{d}{d\theta} h(x,\theta)$ converges uniformly for all θ
                       in a closed subinterval of (a,b)

Then $\frac{d}{d\theta} \sum_{x=0}^{\infty} h(x,\theta) = \sum_{x=0}^{\infty} \frac{d}{d\theta} h(x,\theta)$

Remark: This is a consequence of Lebesgue's D.C.T

Remark: Pointwise convergence:

    For each θ, $\lim_{n\to\infty} \underbrace{\sum_{x=0}^{n} h(x,\theta)}_{S_n(\theta)} = \underbrace{\sum_{x=0}^{\infty} h(x,\theta)}_{S(\theta)}$

Uniform convergence:

    $\sup |S_n(\theta) - S(\theta)| \to 0$ as $n \to \infty$
    $\theta \in (a,b)$

Recall: $A \subset \mathbb{R}$.

    $\sup A$ = smallest $s \in \mathbb{R}$ s.t. $s \geq a \ \forall \ a \in A$

       if $A = [u, v]$, then $\sup A = v$

       if $A = (u, v)$, then $\sup A = v$

Ex (Geometric): Let $X \sim$ Geometric(θ), $0 < \theta < 1$; $P(X=x) = \theta(1-\theta)^x$, $x=1,2,...$

    valid pmf since $\sum_{x=0}^{\infty} \theta(1-\theta)^x = \theta \sum_{x=0}^{\infty} (1-\theta)^x$

                    uses $\sum_{k=0}^{\infty} r^k = \frac{1}{1-r}$, $|r| < 1$

                    and $|1-\theta| < 1$ since $0 < \theta < 1$

                  $= \theta \cdot \frac{1}{1-(1-\theta)} = \theta \frac{1}{\theta} = 1$

    (*) Differentiating both sides wrt θ, assuming we can
       interchange Σ and $\frac{d}{d\theta}$:

       $\sum_{x=0}^{\infty} (1-\theta)^x - \theta x(1-\theta)^{x-1} = 0$

       $\Rightarrow \sum_{x=0}^{\infty} (1-\theta)^x = \sum_{x=0}^{\infty} \theta x (1-\theta)^{x-1}$

       $\underbrace{\frac{1}{1-(1-\theta)}}_{} = \frac{1}{1-\theta} \sum_{x=0}^{\infty} \theta x (1-\theta)^x$

       $\Rightarrow \frac{1}{\theta} = \frac{1}{1-\theta} E(X) \Rightarrow E(X) = \frac{1-\theta}{\theta}$

    Justifying (*),

      • $h(x,\theta) = \theta(1-\theta)^x$ and
      $\forall \ \theta \in (a,b)$ $\sum_{x=0}^{k} h(x,\theta) = \theta \sum_{x=0}^{k} (1-\theta)^x = \theta \frac{1-(1-\theta)^{k+1}}{1-(1-\theta)} \to 1$ as $k \to \infty$

                since $\sum_{x=0}^{k} r^x = \frac{1-r^{k+1}}{1-r}$

      • $\frac{d}{d\theta} h(x,\theta) = (1-\theta)^x - \theta x (1-\theta)^{x-1}$ continuous in θ,
           so h is continuously diff. in θ

      • convergence:
        Pointwise convergence: for each $\theta \in (0,1)$
          $S_k(\theta) = \sum_{x=0}^{k} \frac{d}{d\theta} h(x,\theta) = \sum_{x=0}^{k} (1-\theta)^x - \theta x (1-\theta)^{x-1}$

                  $= \frac{1-(1-\theta)^{k+1}}{\theta} + \theta \sum_{x=0}^{k} \frac{d}{d\theta} (1-\theta)^x$

                  $= \frac{1-(1-\theta)^{k+1}}{\theta} + \theta \frac{d}{d\theta} \sum_{x=0}^{k} (1-\theta)^x$

                  $= \frac{1-(1-\theta)^{k+1}}{\theta} + \theta \frac{d}{d\theta} \frac{1-(1-\theta)^{k+1}}{\theta}$

                  $= \frac{1-(1-\theta)^{k+1}}{\theta} + \theta \frac{(k+1)(1-\theta)^k - (1-(1-\theta)^{k+1})}{\theta^2}$

             by quotient rule: $f(x) = \frac{u(x)}{v(x)}$, $f'(x) = \frac{v(x)u'(x) - u(x)v'(x)}{(v(x))^2}$

                  $= (k+1)(1-\theta)^k$

         So, $S_k(\theta) \to 0$ as $k \to \infty$

      Uniform convergence: Take any $[c,d] \in (0,1)$ and check
             $\sup_{\theta \in [c,d]} |S_k(\theta) - \theta| = \sup_{\theta \in [c,d]} (k+1)(1-\theta)^k \to 0$ as $k \to \infty$. $\square$

Interchanging Integral and Sum

Thm. Suppose $\sum_{x=0}^{\infty} h(x,\theta)$ exists (i.e. converges pointwise) $\forall \ \theta \in [a,b]$

    Moreover, assume

         i) $h(x,\theta)$ is continuous in θ for each fixed x

         2) $\sum_{x=0}^{\infty} h(x,\theta)$ converges uniformly on $[a,b]$

    Then, $\int_a^b \sum_{x=0}^{\infty} h(x,\theta) \ d\theta = \sum_{x=0}^{\infty} \int_a^b h(x,\theta) d\theta$

3.1: Discrete Distributions

<u>Recall</u>: RV X is discrete if the range of X (i.e. $\mathcal{X}$) is countable

---

<u>Def (Uniform RV)</u>:= $X \sim \text{Uniform}(1,N)$ if

$$\boxed{P(X=x) = \frac{1}{N}, \quad x=1,\dots,N.}$$

we sometimes write

$$\boxed{P(X=x \mid N) = \frac{1}{N}}$$

<u>Remarks</u>: $E(X) = \frac{N+1}{2}$

$V(X) = \frac{(N+1)(N-1)}{12} = \frac{(N-1)^2}{12}$

<u>Def (Hypergeometric Dist.)</u>:=

Experiment: N balls, M red, $N-M$ green

Pick K balls uniformly at random (w/o replacement)

Let X: number of reds.

Then,

$$\boxed{P(X=x \mid N,M,K) = \frac{\binom{M}{X}\binom{N-M}{K-X}}{\binom{N}{K}}, \quad x=0,1,\dots,K}$$

<u>remarks</u>: · counting argument implies $\sum_{x=0}^{K} \binom{M}{x}\binom{N-M}{K-x} = \binom{N}{K}$

· $E(X) = \frac{KM}{N}$, $V(X) = \frac{KM}{N} \frac{(N-M)(N-K)}{N(N-1)}$

Ex(Hypergeometric): $E(X) = \sum_{x=0}^{K} x \binom{M}{x}\binom{N-M}{K-x} \frac{1}{\binom{N}{K}}$

<mark>PS</mark> where $x\binom{M}{x} = M\binom{M-1}{x-1}$ and

$\frac{\binom{N-1}{K-1}}{\binom{N}{K}} = \frac{\frac{(N-1)!}{(K-1)!(N-K)!}}{\frac{N!}{K!(N-K)!}} = \frac{(N-1)!}{(K-1)!(N-K)!} \cdot \frac{K!(N-K)!}{N!} = \frac{(N-1)!K!}{(K-1)!N!} = \frac{K}{N}$

$= \sum_{x=1}^{K} M\binom{M-1}{x-1}\binom{N-M}{K-x} \left(\frac{M}{K}\binom{N-1}{K-1}\right)^{-1}$

$= \frac{KM}{N} \sum_{x=1}^{K} \binom{M-1}{x-1}\binom{N-M}{K-x} / \binom{N-1}{K-1}$

letting $y = x-1 \Leftrightarrow x = y+1$

$= \frac{KM}{N} \sum_{y=0}^{K-1} \binom{M-1}{y}\binom{(N-1)-(M-1)}{K-1-y} / \binom{N-1}{K-1}$

$= \frac{KM}{N} \sum_{y=0}^{K-1} P(Y=y \mid N-1, M-1, K-1)$

$= \frac{KM}{N}$

Ex(Hypergeometric): Let $N=25$ items, $M=$ # defective items

we sample K items from the lot, say $K=10$, and find that none are defective.

Q: what can we say about M?

From above, $X \mid N,M,K \sim \text{Hypergeometric}(N,M,K)$

Recall $p(x) = \frac{\binom{M}{x}\binom{N-M}{K-x}}{\binom{N}{K}}$, $x=0,1,\dots,K$.

So, $P(X=0 \mid N=25, M=6, K=10) = \frac{\binom{M}{x}\binom{N-M}{K-0}}{\binom{N}{K}} = \frac{\binom{6}{0}\binom{19}{10}}{\binom{25}{10}} = 0.028.$

$\Rightarrow P(X=0 \mid N=25, K=10, M \geq 6) \leq 0.028$, i.e.

if $M \geq 6$, then the observed event "$X=0$" is highly unlikely

---

<u>Def (Bernoulli RV)</u>:= $X \sim \text{Bernoulli}(p)$, $0 \leq p \leq 1$

if $\begin{cases} P(X=1 \mid p) = p \\ P(X=0 \mid p) = 1-p \end{cases}$

<u>Remarks</u>: $E(X) = 1 \cdot p + 0 \cdot (1-p) = p$

$E(X^2) = E(X) = p$

$V(X) = E(X^2) - (E(X))^2 = p - p^2 = p(1-p)$

<u>Def (Binomial Dist.)</u>:= Perform n-independent-and-identically-distributed (iid) Bernoulli trials.

Let Y: number of successes (i.e. 1's).

Then $Y \sim \text{Binomial}(n,p)$ and

$$\boxed{P(Y=y \mid n,p) = \binom{n}{y} p^y (1-p)^{n-y}}$$

<u>Remarks</u>: $\begin{cases} E(Y) = np \\ V(Y) = np(1-p) \\ M_X(t) = (pe^t + 1-p)^n \end{cases}$

<u>Def (Poisson Dist.)</u>: Let $X \sim \text{Poisson}(\lambda)$, $\lambda > 0$.

$$\boxed{P(X=x \mid \lambda) = e^{-\lambda} \frac{\lambda^x}{x!}, \quad x=0,1,\dots}$$

<u>Remarks</u>: · $e^\lambda = 1 + \lambda + \frac{\lambda^2}{2!} + \dots + \frac{\lambda^x}{x!} + \dots$

· arises in modelling the following (count) experiments:
  + number of buses arriving with a period of time
  + number of stars observed in a skymap
  + number of incidents found in a segment of a network

· $E(X) = \lambda = V(X)$, $M_X(t) = \exp(\lambda(e^t - 1))$.

<u>Ex (Poisson Problem):</u> A call operator handles, on average, 5 calls/3 min.

Find P(no calls in the next minute).

Let X: number of calls in the next minute.

Assume $X \sim \text{Poisson}(\lambda)$, $\lambda = 5/3$.

Recall $p(x) = e^{-\lambda} \frac{\lambda^x}{x!}$

Then $E(X) = \lambda = 5/3$ and

$P(X=0 | \lambda = 5/3) = e^{-5/3} \frac{(5/3)^0}{0!} = e^{-5/3} = 0.189$

<u>Prop (Poisson Approx.)</u> (of the Binomial Dist.):

If $X_n \sim \text{Binomial}(n, p_n)$ and $np_n \to \lambda$ as $n \to \infty$

then $X_n \xrightarrow{d} Y$ where $Y \sim \text{Poisson}(\lambda)$.

<u>Ex (Poisson Approx. to Binomial Dist.):</u> A typesetter, on average, makes one error per 500 words typeset.

A typical page has 300 words.

Find P(≤2 errors in a 5-page essay)

Let X: number of errors in 5-pages

Assume $X \sim \text{Binomial}(n, p)$ where

$p(x) = \binom{n}{x} p^x (1-p)^{n-x}$ and here,

$n = 300 \times 5 = 1500$

$p = 1/500$

Then,

$P(X \le 2 | n = 1500, p = \frac{1}{500}) = \sum_{x=0}^{2} \binom{1500}{x} \left(\frac{1}{500}\right)^x \left(\frac{499}{500}\right)^{1500-x}$

$= 0.4230$

Using the Poisson Approx., i.e. $X \overset{d}{\approx} Y$,

$Y \sim \text{Poisson}(\lambda)$, $\lambda = np = 3$ and $p(y) = e^{-\lambda} \frac{\lambda^y}{y!}$

$\Rightarrow P(X \le 2 | \lambda) \approx P(Y \le 2 | \lambda)$

$= \sum_{y=0}^{2} e^{-\lambda} \frac{\lambda^y}{y!} = e^{-3}(1 + 3 + \frac{3^2}{2!}) = 0.4232$

<u>Def (Negative Binomial Dist.):=</u>

Experiment: (count the number of independent Bernoulli(p)
trials until obtaining r successes

$X \sim \text{Neg Binom}(r, p)$, $0 \le p \le 1$, $r \in \mathbb{N}_+$

$\boxed{P(X=x) = \binom{x-1}{r-1} p^r (1-p)^{x-r}, \quad x = r, r+1, \dots}$

<u>corollary:</u> Let $Y = X - r$ represent the number of failures that occur before
the r-th success in a sequence of independent Bernoulli(p) trials

Then, $\boxed{P(Y=y | p, r) = P(X = y+r | p, r) = \binom{y+r-1}{r-1} p^r (1-p)^y, \quad y = 0, 1, \dots}$

$\overset{(*)}{=} (-1)^y \binom{-r}{y} p^r (1-p)^y$

<u>Remark:</u> (*) Motivates the "negative binomial" label

• Here, $n = y + r - 1 \Rightarrow y = n - r + 1$

• $E(Y) = \frac{r(1-p)}{p}$, $\text{var}(Y) = \frac{r(1-p)}{p^2}$

<u>Prop:</u> Let $\mu = E(Y) = \frac{r(1-p)}{p}$. Then

$\boxed{\text{var}(Y) = \frac{1}{r} \mu^2 + \mu}$

<u>Remark:</u> colloquially the 'quadratic relation'

<u>Prop (Poisson Approx.):</u> If $r \to \infty$, $p \to 1$ such that $r(1-p) \to \lambda$,

then $\begin{cases} E(Y) \to \lambda \\ V(Y) \to \lambda \end{cases}$

Moreover, $Y \xrightarrow{d} \text{Poisson}(\lambda)$

<u>Def (Geometric Dist.):=</u> Let $X \sim \text{Geometric}(p)$, $p \in (0, 1)$.

$\boxed{P(X=x | p) = p(1-p)^{x-1}, \quad x = 1, 2, \dots}$

<u>Remarks:</u> this is a special case of NegBinom$(p, r=1)$;

thus $\begin{cases} E(X) = (1-p)/p \\ V(X) = (1-p)/p^2 \end{cases}$

<u>Prop ('memoryless property'):</u> If $s > t$, then

$\boxed{P(X > s | X > t) = P(X > s - t)}$

<u>Remarks:</u> Given $X > t$ (no success in the first t trials), the dist. of additional waiting time
until the first success does not depend on t, i.e.
past failures do not change the tail prob's for the future