Lecture 9 - 9/12

.Review

① Why we care about distributions?

② Common pmfs:

1. Bernoulli distribution: $Ber(p)$, $P_X(1)=p$, $P_X(0)=1-p$
2. Binomial distribution: $Bi(n,p)$, $P_X(k)=\binom{n}{k}p^k(1-p)^{n-k}$ $k\in\{0,1,\dots,n\}$
3. Geometric distribution: $Geo(p)$, $P_X(k)=(1-p)^{k-1}p$ $k\in\{1,2,\dots\}$
4. Negative Binomial distribution. $NB(r,p)$, $P_X(n)=\binom{n-1}{r-1}p^r(1-p)^{n-r}$, $n\in\{0,1,2,\dots\}$
5. Poisson distribution: $Poi(\lambda)$, $P_X(k)=e^{-\lambda}\frac{\lambda^k}{k!}$, $k\in\{0,1,2,\dots\}$

Today:

1. Important remark
2. Memoryless property
3. Poisson distribution
4. Maximum likelihood estimator
5. Markov inequality

1. important remark:

For defining a pmf, make sure you specify support. If you write a formula for $P_X(k)$ make sure to mention domain of k.

2. Memoryless property of $Geo(p)$:

Suppose X is a $Geo(p)$ random variable, and m,n are positive integers.

$$P(X>m+n \mid X>n) = \frac{P(\{X>m+n\}\cup\{X>n\})}{P\{X>n\}}$$

$$= \frac{(1-p)^{m+n}}{(1-p)^n} = (1-p)^m = P(X>m)$$

Suppose an accident happens in green street with prob. p each day independent of each other. Suppose that accident has not happened in the last 100 days $(n=100)$, asking "what is the probability that there is going to be no accident for the next 10 days as well" is same as asking "what is the probability that there is going to be no accident for the next 10 days starting from now"!

no accident for the next 10 days as well is same as asking what is the probability that there is going to be no accident for the next 10 days starting from now"!

3. Poisson distribution.

Suppose that X is a $Bi(n,p)$ random variable where $np=\lambda$.

$$P_X(k) = \binom{n}{k} p^k (1-p)^{n-k} = \frac{n(n-1)\cdots(n-k+1)}{k!} p^k (1-p)^{n-k}$$

$$= \frac{\lambda^k}{k!} \left(\frac{n(n-1)\cdots(n-k+1)}{n^k}\right)\left(1-\frac{\lambda}{n}\right)^{n-k}$$

If λ is fixed, k is fixed and $n \longrightarrow \infty$ (which means $p \to 0$ as np is fixed)

$$P_X(k) \longrightarrow \frac{\lambda^k}{k!} \cdot 1 \cdot e^{-\lambda}$$

Def: We say a random variable has $Poi(\lambda)$ distribution if:

$$P_X(k) = e^{-\lambda} \frac{\lambda^k}{k!} \qquad k \in \{0,1,2,\ldots\}$$

- We need to make sure $P_X(k)$ satisfies axioms of probability:

$$\sum_{k=0}^{\infty} P_X(k) = \sum_{k=0}^{\infty} e^{-\lambda} \frac{\lambda^k}{k!} = 1.$$

Recall that $e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!}$, hence, $\sum_{k=0}^{\infty} \frac{\lambda^k}{k!} = e^{\lambda}$.

$$E[X] = \sum_{k=0}^{\infty} k \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} \frac{e^{-\lambda}\lambda^k}{(k-1)!} = \lambda \sum_{\ell=0}^{\infty} \frac{e^{-\lambda}\lambda^\ell}{\ell!} = \lambda$$

This should have been expected as $Poi(\lambda)$ is an approximation of $Bi(n,p)$ with $\lambda = np$, and expected value of $Bi(n,p)$ is np.

Using some ideas.

$$E[X^2] = \sum_{k=0}^{\infty} k^2 \frac{e^{-\lambda}\lambda^k}{k!} = \sum_{k=1}^{\infty} k \frac{e^{-\lambda}\lambda^k}{(k-1)!} = \sum_{\ell=0}^{\infty} (\ell+1) \frac{e^{-\lambda}\lambda^{\ell+1}}{\ell!} = \lambda^2 + \lambda$$

Hence, variance of poisson distribution is $Var(X) = E[X^2] - (E[X])^2 = \lambda$. Notice that for Binomial distribution $Var(X) = np(1-p) = np - (np)p$. Fixing $\lambda = np$, taking $n \to \infty$, $p \to 0$, we get the result.

note: Using comparison between poisson & binomial is not a rigorous proof. To be precise, we are changing order of limit & infinite sum which needs justification & is beyond the scope of this class.

4. Maximum likelihood estimator.

Consider a family of random variables parametrized with θ:

e.g. family of $Bi(n,\theta)$ random variables, $\theta \in (0,1)$, for each θ denote its pmf with $p_\theta(k) = \binom{n}{k} \theta^k (1-\theta)^{n-k}$

e.g. family of $Geo(\theta)$ random variables, $\theta (0,1)$, for each θ denote its pmf with $p_\theta(k) = (1-\theta)^{k-1}\theta$.

suppose that we know random variable X belongs to some family that is paramctrized θ, but we don't know the value of θ. Suppose that, after running an experiment, we observe a realization $X=k$. Suppose that the pmf of X is given by $p_{\theta^*}(\cdot)$, i.e., $p_X(k) = p_{\theta^*}(k)$.

- $p_{\theta^*}(k)$ is the likelihood of observing $X=k$.

Now, given $X=k$, we want to estimate θ^*. Maximum-likelihood estimator suggest picking $\hat{\theta}_{ML}$ that maximizes probability of observing $X=k$, i.e.,

$$\hat{\theta}_{ML} = \arg\max_{\theta} p_\theta(k)$$

example: Suppose that X has $B(n,p)$ distribution. We know n, but we don't know p. Suppose that we are given a realization $X=k$. What is the ML-estimator of X?

$$p_X(k) = \binom{n}{k} p^k (1-p)^{n-k}$$

$\hat{p}_{ML}$ maximizes $p^k(1-p)^{n-k}$. See example in book!

---

5. Markov inequality.

One of the main statistical quantities associated with random variables is their tail: $P(X \geq k)$.

e.g. number of customers that appear in a bank from 8am to 8:30am is a Poisson random variable with parameter λ. Eeach clerk can service one customer during this interval. How many clerk should work during this time so that with probability 0.9, no one will wait for service?

Let k denote number of clerks and X denote number of customers. We need pick k such that

$$P(X \geq k) \leq 0.1$$

Hence, having easy bounds for tail of random variables is desirable. One such bound is Markov inequality.

Suppose that X is a non-negative random variable. Pick some constant $c > 0$:

$$E[X] = \sum_{u \geq 0} u\, p_X(u) \geq \sum_{u \geq c} u\, p_X(u) \geq \sum_{u \geq c} c\, p_X(u) = c\, P(X \geq c)$$

$$\Rightarrow P(X \geq c) \leq \frac{E[X]}{c}$$