دانشکده مهندسی کامپیوتر
دانشگاه صنعتی شریف

استاد درس: دکتر حمیدرضا ربیعی

بهار ۱۴۰۰

تمرین در خانه سوم

درس یادگیری ماشین آماری

امیر پورمند
شماره دانشجویی: ۹۹۲۱۰۲۵۹

سری سوم تمرین ها

---

۱ سوال ۱

۱.۱ a

پس با توجه به جمله اول سوال در واقع $N(t) = 1$ می باشد. باید توزیع زمان رخداد رو بدست بیاریم.
در ابتدا در نظر میگیریم که متغیر تصادفی s در واقع همان زمان رخدادن ماست.
مثل همیشه برای بدست آوردن توزیع ابتدا CDF آن را محاسبه می‌کنیم. بنابراین داریم

$$\begin{aligned} F_{S|N(t)=0}(s) &= P(S \leq s|N_t = 0) \\ &= P(N(s) = 1|N(t) = 1) \\ &= \frac{P(N(s) = 1, N(t) = 1)}{P(N(t) = 1)} \\ &= \frac{P(N(s) = 1, N(t) - N(s) = 0)}{P(N(t) = 1)} \\ &= \frac{\exp(-\lambda s)(\lambda s)^1 \exp(-\lambda(t-s))((\lambda(t-s))^0/(0!))}{\exp(-\lambda t)(\lambda t)^1/(1!)} \\ &= \frac{exp(-\lambda t)\lambda s}{exp(-\lambda t)(\lambda t)} \\ &= \frac{s}{t} \end{aligned} \tag{۱}$$

بنابراین یک توزیع یکنواخت داریم. البته دقت داشته باشیم متغیر ما در اینجا s است یعنی CDF در واقع به این صورت است.

$$\frac{s-0}{t-0}$$

که این بدین معناست که در بازه ۰ تا t متغیر ما توزیع یکنواخت دارد

۲.۱ b

خب اول کمی mapping انجام دهیم. میدانیم دو تا اتفاق افتاده که آن ها را با s_1, s_2 نشان می دهیم. و مشخصا میدانیم که $0 \leq S_1 \leq S_2 \leq 3$ است
خب حالا باید ابتدا توزیع شرطی s۱ به شرط $N(t) = 2$ را بدست آوریم که برابر است با

$$
\begin{aligned}
F_{S1|N(t)=2}(s1) &= \frac{P(S_1 \leq s1, N(t)=2)}{P(N(t)=2)} \\
&= \frac{P(N(s1) \geq 1, N(t)=2)}{P(N(t)=2)} \\
&= \frac{P(N(s_1)=1, N(t)=2) + P(N(s_1)=2, N(t)=2)}{P(N(t)=2)} \\
&= \frac{P(N(s_1)=1, N(t)-N(s_1)=1) + P(N(s_1)=2, N(t)-N(s1)=0)}{P(N(t)=2)} \\
&= \frac{e^{-\lambda s_1}(\lambda s_1)e^{-\lambda(t-s_1)}\lambda(t-s_1) + e^{-\lambda s_1}(\lambda s_1)^2/2! * e^{-\lambda(t-s_1)}}{e^{-\lambda t}(\lambda t)^2/(2!)} \\
&= \frac{\lambda^2 e^{-\lambda t} s_1(t-s1) + e^{-\lambda t}(\lambda s_1)/2}{\lambda^2 t^2 e^{-\lambda t}/2} \\
&= \frac{2s_1 t - s_1^2}{t^2}
\end{aligned} \tag{۲}
$$

پس داریم توزیع متغیر s۱ به این صورت خواهد بود و از مشتق CDF نسبت به s۱ بدست خواهد آمد.

$$
\begin{aligned}
f_{S1|N(t)=2}(s1) &= \frac{\partial F}{\partial s_1} \\
&= \frac{\partial 2s_1 t - s_1^2}{t^2 \partial s_1} \\
&= \frac{\partial 2s_1 t}{t^2 \partial s_1} - \frac{\partial s_1^2}{t^2 \partial s_1} \\
&= \frac{2t}{t^2} - \frac{2s_1}{t^2} \\
&= \frac{2(t-s_1)}{t^2}
\end{aligned} \tag{۳}
$$

حال باید امیدریاضی توزیع شرطی متغیر s را بدست اوریم که میشود

$$
\mathbb{E}(s_1|N(t)=2) = \int_0^t s_1 \frac{2(t-s_1)}{t^2} ds_1 = \frac{t}{3} \tag{۴}
$$

حال به قسمت دوم سوال میرسیم و باید در اینجا توزیع s_2 را بدست بیاوریم. باز هم به سراغ CDF رفته و بعد از آن از آن مشتق میگیریم تا توزیع شرطی بدست اید. خب شروع میکنیم

$$
\begin{aligned}
F_{S_2|N(t)=2}(s_2) &= \frac{P(S_2 \leq s_2, N(t)=2)}{P(N(t)=2)} \\
&= \frac{P(N(s_2)=2, N(t)=2}{P(N(t)=2} \\
&= \frac{P(N(s_2)=2, N(t)-N(s_2)=0)}{P(N(t)=2)} \\
&= \frac{\frac{e^{-\lambda s_2}(\lambda s_2)^2}{(2!)} e^{-\lambda(t-s_2)}}{\frac{e^{-\lambda t}(\lambda t)^2/(2!)}{2!}} \\
&= \frac{\lambda^2 s_2^2}{\lambda^2 t^2} \\
&= \frac{s_2^2}{t^2}
\end{aligned} \tag{۵}
$$

حال باید طبق روال قبل توزیع شرطی این متغیر را بدست آوریم که برابر مقدار زیر خواهد بود

$$
f_{S2|N(t)=2}(s_2) = \frac{\partial s_2^2}{t^2 \partial s_2} = \frac{2s_2}{t^2} \tag{۶}
$$

و امیدریاضی شرطی متغیر s_2 برابر مقدار زیر خواهد شد

$$
\mathbb{E}[s_2|N(t)=2] = \int_0^t s_2 \frac{2s_2}{t^2} ds_2 = \frac{2}{t^2}\frac{t^3}{3} = \frac{2t}{3} \tag{۷}
$$

۲ سوال ۲

۱.۲ توزیع T۱

خب ابتدا باید توزیع تجمعی T_1 و بعد T_2 را بدست آوریم و بعد مشتق بگیریم که توزیع بدست آید. برای T_1 داریم:

$$
\begin{aligned}
F_{T_1}(t) = P(T_1 \leq t) &= 1 - P(T_1 \geq t) \\
&= 1 - P(N(t)=0) \\
&= 1 - exp(-\int_0^t \frac{1}{1+k} dk) \\
&= 1 - exp(-(ln(t+1) - ln(0+1)) \\
&= 1 - exp(ln(\frac{1}{1+t})) \\
&= 1 - \frac{1}{1+t}
\end{aligned} \tag{۸}
$$

حال تابع توزیع از مشتق نسبت به t بدست می آید.

$$
\begin{aligned}
f_{T_1}(t) &= \frac{\partial}{\partial t} F_{T_1}(t) \\
&= \frac{\partial}{\partial t} \frac{t}{t+1} \\
&= \frac{1(t+1) - 1(t)}{(1+t)^2} \\
&= \frac{1}{(1+t)^2}
\end{aligned} \tag{۹}
$$

۲.۲ توزیع T۲

حال باید توزیع متغیر T_2 را بدست آوریم.

$$
\begin{aligned}
P(T_2 \geq t) &= P(N(t) = 1) + P(N(t) = 0) \\
&= exp(-\int_0^t \frac{1}{1+k} dk)(\int_0^t \frac{1}{1+k} dk) + exp(-\int_0^t \frac{1}{1+k} dk) \\
&= exp(-log(1+t))(log(1+t)) + exp(-log(1+t)) \\
&= \frac{log(1+t)}{1+t} + \frac{1}{1+t}
\end{aligned} \tag{۱۰}
$$

بنابراین داریم

$$
F_{T_2}(t) = P(T_2 \leq t) = 1 - \frac{log(1+t)}{1+t} - \frac{1}{1+t} \tag{۱۱}
$$

که اگر از ان مشتق بگیریم داریم:

$$
f_{T2}(t) = \frac{\partial}{\partial t} F_{T_2}(t) = \frac{ln(t+1)}{(t+1)^2} \tag{۱۲}
$$

البته این زمان رسیدن است و چیزی که در سوال خواسته شده interarrival تایم هست که آن را در ادامه بدست می اورم. یعنی تفاوت در این است که الان T_1 در واقع زمان رسیدن است و در زیر T_1 زمان بین دو اتفاق است که در اینصورت جواب سوال در قسمت زیر می آید.

ابتدا میدانیم $N(T_1 + t) - N(T_1)$ از توزیع پواسون پیروی میکند و نرخ آن برابر خواهد بود با:

$$
\int_{T_1}^{T_1+t} \frac{1}{1+k} dk = ln(1 + T_1 + t) - ln(1 + T_1) = ln\frac{1 + T_1 + t}{1 + T_1} \tag{۱۳}
$$

حال توزیع شرطی T۲ به شرط T۱ برابر خواهد بود با:

$$
\begin{aligned}
P(T_2 > t|T_1) &= P(N(T_1+t) - N(T_1) = 0|T_1) \\
&= exp(-ln\frac{1+T_1+t}{1+T_1}) \\
&= \frac{T_1+1}{T_1+t+1}
\end{aligned} \tag{۱۴}
$$

حالا توزیع T_1 را داریم میتوانیم با ضرب کردن توزیع جوینت را بدست آوریم و سپس مارجینالایز بکنیم و خود احتمال را بدست اوریم.

$$
\begin{aligned}
P(T_2 > t) &= P(T_2 > t|T_1) f_{T_1}(k) \\
&= \int_0^\infty \frac{k+1}{k+t+1}\frac{1}{(1+k)^2}dk \\
&= \int_0^\infty \frac{1}{k+t+1}\frac{1}{1+k}dk \\
&= \frac{ln(1+t)}{t}
\end{aligned} \tag{۱۵}
$$

اکثر راه را رفتیم پس توزیع CDF برابر است با:

$$
F_{T_2}(t) = 1 - P(T_2 > t) = 1 - \frac{ln(1+t)}{t} \tag{۱۶}
$$

فقط یگ مشتق مانده که بگیریم و تمام شود.

$$
f_{T_2}(t) = \frac{\partial}{\partial t}F_{T_2}(t) = \frac{\partial}{\partial t}[1 - \frac{ln(1+t)}{t}] = \frac{\ln(t+1)}{t^2} - \frac{1}{t\,(t+1)} \tag{۱۷}
$$

۳ سوال ۳

۱.۳ a

خب اول میریم سراغ تعریف های ریاضی و فهم بهتر مسئله. پس داریم:

$$N \sim Pois(\lambda) \tag{۱۸}$$

$$S|N \sim Binomial(N, p) \tag{۱۹}$$

$$T|N \sim Binomial(N, 1-p) \tag{۲۰}$$

توزیع N که مشخص است. ابتدا توزیع S را دقیق بدست آوریم. (البته در ابتدا فرض کنیم n را میدانیم!) پس داریم:

$$
P(S = s|N = n) = \binom{n}{s} p^s (1-p)^{n-s} \tag{۲۱}
$$

حال میتوان توزیع S را با قانون احتمال کل بدست آورد!

$$\begin{aligned}
P(S=s) &= \sum_{n=s}^{\infty} P(S=s|N=n)P(N=n) \\
&= \sum_{n=s}^{\infty} \binom{n}{s} p^s (1-p)^{n-s} \frac{e^{-\lambda}\lambda^n}{n!} \\
&= e^{-\lambda} p^s \sum_{n=s}^{\infty} \frac{n!}{(n-s)!s!} \frac{(1-p)^{n-s}\lambda^n}{n!} \\
&= \frac{e^{-\lambda}p^s}{s!} \sum_{n=s}^{\infty} \frac{(1-p)^{n-s}\lambda^n}{(n-s)!} \\
&= \frac{e^{-\lambda}p^s}{s!} \sum_{n=0}^{\infty} \frac{(1-p)^{n}\lambda^{n+s}}{n!} \\
&= \frac{e^{-\lambda}\lambda^s p^s e^{(1-p)\lambda}}{s!} \\
&= \frac{e^{-p\lambda}}{(p\lambda)^s} s! \\
&\sim Pois(p\lambda)
\end{aligned} \tag{۲۲}$$

برای حل بالا از بسط تیلور یا سری توانی استفاده شده است که به شکل زیر است:

$$e^x = \sum_{k=0}^{\infty} \frac{x^k}{k!} \tag{۲۳}$$

به طرز مشابه میتوان نشان داد که $T \sim Pois((1-p)\lambda)$

۲.۳ b

برای نشان داده این که دو متغیر از نظر آماری مستقل هستند باید نشان دهیم که توزیع جوینت آنها معادل ضرب توزیع حاشیه ای آنها است. پس داریم:

$$\begin{aligned} f_{S,T}(s,t) &= P(S=s, T=t) \\ &= P(S=s, N=t+s) \\ &= P(S=s|N=t+s)P(N=t+s) \\ &= \binom{t+s}{s} p^s(1-p)^t \frac{e^{-\lambda}(\lambda)^{t+s}}{(t+s)!} \\ &= \frac{(t+s)!}{t!s!} p^s(1-p)^t e^{-\lambda}\lambda^t\lambda^s \frac{1}{(t+s)!} \\ &= \frac{e^{-\lambda(1-p+p)}p^s\lambda^s(1-p)^t\lambda^t}{t!s!} \\ &= \frac{e^{-\lambda(1-p)}((1-p)\lambda)^t}{t!} \frac{e^{-\lambda p}(p\lambda)^s}{s!} \\ &= P(T=t)P(S=s) = Solved! \end{aligned} \tag{۲۴}$$

۳.۳ c

خب فرض میکنیم متغیر تصادفی N یک عدد ثابت مثل n باشد. پس داریم:

$$\begin{aligned} f_{S,T}(s,t) = P(S=s, T=t) = P(S=s, N=s+t) \\ = P(N=s+t)P(S=s|N=s+t) = P(S=s) \end{aligned} \tag{۲۵}$$

پس در این حالت می توان گفت هر کدام از S و T به یکدیگر کاملا وابسته هستند.