SELECTED PROOFS

DeMorgan's formulas:
The first one is clear from Venn diagram, or the following 'truth' table:

A	B	$A \cup B$	$\overline{A \cup B}$	$\bar{A}$	$\bar{B}$	$\bar{A} \cap \bar{B}$
T	T	T	F	F	F	F
T	F	T	F	F	T	F
F	T	T	F	T	F	F
F	F	F	T	T	T	T

The second one can be derived from the first by changing A to $\bar{A}$ and B to $\bar{B}$, thus:

$$\overline{\bar{A} \cup \bar{B}} = A \cap B$$

(since $\overline{\overline{A}} = A$), and taking complement of each side:

$$\bar{A} \cup \bar{B} = \overline{A \cap B}$$

Product rule:

$$\begin{aligned}
&\Pr(A) \times \Pr(B \mid A) \times \Pr(C \mid A \cap B) = \\
&\Pr(A) \times \frac{\Pr(A \cap B)}{\Pr(A)} \times \frac{\Pr(A \cap B \cap C)}{\Pr(A \cap B)} = \\
&\Pr(A \cap B \cap C)
\end{aligned}$$

Total probability formula:

$$\begin{aligned}
&\sum_{\text{All } k} \Pr(B \mid A_k) \Pr(A_k) = \\
&\sum_{\text{All } k} \frac{\Pr(A_k \cap B)}{\Pr(A_k)} \Pr(A_k) = \\
&\sum_{\text{All } k} \Pr(A_k \cap B) = \text{ (because they are disjoint)} \\
&\Pr\left[\cup_{\text{All } k}(A_k \cap B)\right] = \text{ (by distributive law)} \\
&\Pr\left[(\cup_{\text{All } k} A_k) \cap B\right] = \Pr(B) \text{ (since } \cup_{\text{All } k} A_k = \Omega)
\end{aligned}$$

'Total mean' formula:

$$\begin{aligned}
\mathbb{E}(X) = \sum_{\text{All } i} i \times \Pr(X &= i) \text{ (definition of expacted value)} \\
= \sum_{\text{All } i} i \times \sum_{\text{All } k} \Pr(X &= i \mid A_k) \Pr(A_k) \text{ (by the previous formula)} \\
= \sum_{\text{All } k} \sum_{\text{All } i} i \times \Pr(X &= i \mid A_k) \Pr(A_k) \text{ (interchange the 2 summations)} \\
= \sum_{\text{All } k} \mathbb{E}(X &= i \mid A_k) \Pr(A_k) \text{ (by definition of conditional expacted value)}
\end{aligned}$$

$-1 \leq \rho \leq 1$ **(or, equivalently, $\rho^2 \leq 1$):**

$$\begin{aligned}
&\mathbb{E}\{[X-\mu_x-\lambda(Y-\mu_y)]^2\} = \\
&\mathbb{E}[(X-\mu_x)^2] - 2\lambda\mathbb{E}[(X-\mu_x)(Y-\mu_y)] + \lambda^2\mathbb{E}[(Y-\mu_y)^2] = \\
&\mathrm{Var}(X) - 2\lambda\mathrm{Cov}(X,Y) + \lambda^2\mathrm{Var}(Y) \geq 0
\end{aligned}$$

for any λ (averaging non-negative quantity yields non-negative answer).

The last expression will have the smallest possible value when the λ-derivative is zero, namely

$$-2\mathrm{Cov}(X,Y) + 2\lambda\mathrm{Var}(Y) = 0$$

or

$$\lambda = \frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}$$

Substituting this into the same expression yields

$$\begin{aligned}
&\mathrm{Var}(X) - 2\frac{\mathrm{Cov}(X,Y)}{\mathrm{Var}(Y)}\mathrm{Cov}(X,Y) + \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(Y)^2}\mathrm{Var}(Y) = \\
&\mathrm{Var}(X) - \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(Y)} \geq 0
\end{aligned}$$

This implies

$$1 - \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(X)\mathrm{Var}(Y)} \geq 0$$

or

$$1 \geq \frac{\mathrm{Cov}(X,Y)^2}{\mathrm{Var}(X)\mathrm{Var}(Y)} = \rho^2$$

Expected value of linear combination of RVs:

$$\begin{aligned}
&\mathbb{E}\left(aX+bY+c\right) = \int\limits_{\text{All } x \text{ and } y} (ax+by+c)\cdot f(x,y)dxdy = \\
&a\cdot\int\limits_{\text{All } x \text{ and } y} x\cdot f(x,y)dxdy + b\cdot\int\limits_{\text{All } x \text{ and } y} x\cdot f(x,y)dxdy + c = \\
&a\mathbb{E}\left(X\right) + b\mathbb{E}(Y) + c
\end{aligned}$$

(since an integral is a linear operator, which means: a constant can be taken out, and: integrating a sum can be done by integrating the terms individually and adding the answers). In a discrete case, we use summation instead of integration (rest is the same).

Variance of linear combination of RVs:

$$\mathrm{Var}\,(aX + bY + c) =$$
$$\int\limits_{\text{All } x \text{ and } y} [(ax + by + c) - (a\mu_x + b\mu_y + c)]^2 \cdot f(x, y)dxdy =$$
$$\int\limits_{\text{All } x \text{ and } y} [a(x - \mu_x) + b(y - \mu_y)]^2 \cdot f(x, y)dxdy =$$
$$\int\limits_{\text{All } x \text{ and } y} [a^2(x - \mu_x)^2 + b^2(y - \mu_y)^2 + 2ab(x - \mu_x)(y - \mu_y)] \cdot f(x, y)dxdy =$$
$$a^2\mathrm{Var}(X) + b^2\mathrm{Var}(Y) + 2ab\mathrm{Cov}((X, Y)$$

Properties of MGF:
Since:

$$M_X(u) = \mathbb{E}[e^{uX}] = \int\limits_{\text{All } x} e^{ux} f(x)dx$$

then, quite clearly

$$\begin{aligned} M_{aX+b}(u) &= \mathbb{E}[e^{u(aX+b)}] = \int\limits_{\text{All } x} e^{aux+bu} f(x)dx = \\ e^{bu} \int\limits_{\text{All } x} e^{aux} f(x)dx &= e^{bu} \cdot M_X(au) \end{aligned}$$

When X and Y are independent, then

$$M_{X+Y}(u) = \mathbb{E}[e^{u(X+Y)}] =$$
$$\int\limits_{\text{All } y} \int\limits_{\text{All } x} e^{ux+uy} f_X(x) \cdot f_Y(y)dxdy = \text{(the integral is separable)}$$
$$\int\limits_{\text{All } x} e^{ux} f_X(x)dx \int\limits_{\text{All } y} e^{uy} f_Y(y)dy = M_X(u) \cdot M_Y(y)$$

Finally:

$$\begin{aligned} M_X(u) &= \mathbb{E}[e^{uX}] = \mathbb{E}[1 + uX + \frac{u^2}{2}X^2 + \frac{u^3}{3!}X^3 + ...] = \\ & 1 + u\mathbb{E}[X] + \frac{u^2}{2!}\mathbb{E}[X^2] + \frac{u^3}{3!}\mathbb{E}[X^3] + ... \end{aligned}$$

which proves that the simple moments are coefficients in the Taylor expansion of $M_X(u)$.

Properties of PGF:
Since:

$$P_X(z) = \mathbb{E}[z^X] = \sum_{\text{All } i} z^i f(i)$$

then, when X and Y are independent, we get

$$M_{X+Y}(z) = \mathbb{E}[z^{(X+Y)}] =$$
$$\sum_{\text{All } j} \sum_{\text{All } i} z^{i+j} f_X(i) \cdot f_Y(j) =$$
$$\sum_{\text{All } i} z^i f_X(i) \sum_{\text{All } y} z^j f_Y(j) = P_X(z) \cdot P_Y(z)$$

Since

$$P_X^{(k)}(z) = \sum_{\text{All } i} i(i-1)...(i-k+1) z^{i-k} f(i)$$

we get

$$P_X^{(k)}(z)\Big|_{z=1} = \sum_{\text{All } i} i(i-1)...(i-k+1) f(i) = \mathbb{E}\left[X(X-1)....(X-k+1)\right]$$

i.e. the kth factorial moment.

Similarly

$$P_X^{(k)}(z)\Big|_{z=0} = k! f(k)$$

Convolution:

Since

$$\Pr(X+Y<v) = \int_{-\infty}^{\infty} \int_{-\infty}^{v-x} f(x,y) dy dx$$

The pdf of $V = X+Y$ is the v derivative of the above, namely

$$\int_{-\infty}^{\infty} f(x, v-x) dx$$

Here, we need to recall that

$$\frac{d}{dv} \int_{-\infty}^{g(v)} f(y) dy = g'(v) \cdot f[g(v)]$$

in general.

Central Limit theorem:

We need the MGF of $\frac{\bar{X}-\mu}{\sigma/\sqrt{n}} = \sum_{i=1}^{n} \frac{X_i-\mu}{\sigma\sqrt{n}}$. The MGF of each $\frac{X_i-\mu}{\sigma\sqrt{n}}$ expands to:

$$1 + \frac{u^2}{2n} + \frac{u^3}{3!} \frac{\mathbb{E}[(X-\mu)^3]}{\sigma^3 n^{3/2}} + ...$$

Raising this to the power of n, and taking the $n \to \infty$ limit yields

$$\left[1 + \frac{u^2}{2n} + \frac{u^3}{3!} \frac{\mathbb{E}[(X-\mu)^3]}{\sigma^3 n^{3/2}} + ...\right]^n \underset{n\to\infty}{\to} e^{u^2/2}$$

since terms with a higher-than-1 power of n in the denomination don't matter.

This is the MGF of the standardized Normal distribution with the pdf equal to

$$f(z) = \frac{e^{-z^2/2}}{\sqrt{2\pi}}$$

(for all real z).

Verification:

$$\frac{1}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{uz} e^{-z^2/2} dz = \frac{e^{u^2/2}}{\sqrt{2\pi}} \int_{-\infty}^{\infty} e^{-(z-u)^2/2} dz = e^{u^2/2}$$

Composition:

$$\mathbb{E}\left(z^{S_N}\right) = \sum_{n=0}^{\infty} \mathbb{E}\left(z^{S_N} \mid N = n\right) \Pr(N = n) =$$
$$\sum_{n=0}^{\infty} \mathbb{E}\left(z^{S_n}\right) \Pr(N = n) = \sum_{n=0}^{\infty} P_X(z)^n \Pr(N = n) =$$
$$\sum_{n=0}^{\infty} \mathcal{Z}^n \Pr(N = n) = P_N(\mathcal{Z}) = P_N[P_X(z)]$$

where

$$S_N = \sum_{i=1}^{N} X_i$$

N is a RV with PGF given by $P_N(z)$, and the X_i are IID from a distribution with the following PGF:

$$P_X(z) \equiv \mathcal{Z}$$

Binomial distribution:

The sample space of the experiment consists of all n-letter 'words' made up of two letters, S and F (success and failure) - we know there are 2^n of them. Each of these has the probability of

$$p^i (1-p)^{n-i}$$

where p is the probability of a single success, and i is the number of S letters the 'word' contains. We also know that there are $\binom{n}{i}$ 'words' with i letters S, the total probability of getting i successes (in any order) is thus

$$\Pr(X = i) = \binom{n}{i} p^i (1-p)^{n-i}$$

where i ranges from 0 to n inclusive.

Note that proving the Binomial Theorem, which states that

$$\sum_{i=0}^{n} \binom{n}{i} A^i B^{n-i} = (A+B)^n$$

for any A and B, would be similar: expand $(A+B)\cdot(A+B)\cdot\ldots\cdot(A+B)$ using the distributive law, and get a sum of all 'words' consisting of A and B, etc.

With the help of the Binomial Theorem, the PGF of X is

$$P(z) \equiv \sum_{i=0}^{n} \binom{n}{i} (pz)^i (1-p)^{n-i} = (1-p+pz)^n$$

The corresponding mean is

$$\mu = n(1-p+pz)^{n-1}p\Big|_{z=1} = np$$

the second factorial moment is

$$n(n-1)(1-p+pz)^{n-2}p^2\Big|_{z=1} = n(n-1)p^2$$

implying, for the variance:

$$\sigma^2 = n(n-1)p^2 + \mu - \mu^2 = np(1-p)$$

Geometric:

It's obvious that the probability of $i-1$ failures followed by a success is

$$\Pr(X=i) = pq^{i-1}$$

for any positive (integer) i, where $q \equiv 1-p$.

The corresponding PGF is:

$$P(z) \equiv \sum_{i=1}^{\infty} pq^{i-1}z^i = pz(1+qz+q^2z^2+q^3z^3+.....) = \frac{pz}{1-qz}$$

since

$$1+A+A^2+A^3+.... = \frac{1}{1-A}$$

for any $|A|<1$ - to prove that, do the Taylor expansion of $\frac{1}{1-A}$ (as a function of A).

Expanding PGF in terms of z at 1 (Maple is quite good at this) yields

$$\frac{pz}{1-(1-p)z} \simeq 1 + \frac{z-1}{p} + \frac{(1-p)}{p^2}(z-1)^2 + ...$$

which implies that the mean is $\frac{1}{p}$ and the variance is $\frac{2(1-p)}{p^2} + \frac{1}{p} - \frac{1}{p^2} = \frac{1-p}{p^2}$.

Negative binomial:

Since it's a sum of k independent RVs of geometric type, its mean and variance are k times the previous two results, and

$$P(z) = \left(\frac{pz}{1-qz}\right)^k$$

To get the kth success at the ith trial, the first $i-1$ trials must result in exactly $k-1$ successes (in any order), and the ith trial must be a success. Thus, we get

$$\Pr(X=i) = \binom{i-1}{k-1} p^{k-1} q^{i-1} \cdot p = p^k \binom{i-1}{i-k} q^{i-1}$$

where i is a positive integer $\geq k$.

Poisson:

It can be introduced as a limit of the Binomial distribution, when $n \to \infty$ but the mean is kept constant at λ (this implies that $p = \frac{\lambda}{n}$), namely:

$$\Pr(X=i) = \lim_{n\to\infty} \frac{n(n-1)(n-2)...(n-i+1)}{i!} \frac{\lambda^i}{n^i} (1-\frac{\lambda}{n})^n = \frac{\lambda^i}{i!} e^{-\lambda}$$

since $\frac{n-1}{n}$, $\frac{n-2}{n}$, ... all tend to 1, and

$$\lim_{n\to\infty} (1+\frac{a}{n})^n = e^a$$

for any a.

Its PGF is

$$P(z) = e^{-\lambda} \sum_{i=0}^{\infty} \frac{(\lambda z)^i}{i!} = e^{-\lambda(1-z)} \simeq 1 + \lambda(z-1) + \frac{\lambda^2 (z-1)^2}{2} + ...$$

implying that the mean is λ and the variance is $\lambda^2 + \lambda - \lambda^2 = \lambda$.

Exponential:

It can be introduced as a limit of the geometric distribution, when we perform n trials every unit of time, keeping the mean time (of getting the first success) fixed at β (this implies that $p = \frac{1}{n\beta}$), thus:

$$F(x) = \Pr(X \leq x) = \lim_{n\to\infty} 1 - (1 - \tfrac{1}{n\beta})^{nx} = 1 - e^{-x/\beta}$$

for any $x > 0$. This implies that

$$f(x) = F'(x) = \tfrac{1}{\beta} e^{-x/\beta}$$

and the MGF is

$$\tfrac{1}{\beta} \int_0^{\infty} e^{-x/\beta + xu} dx = \tfrac{1}{\beta} \left. \frac{e^{-x/\beta+xu}}{u - \frac{1}{\beta}} \right|_{x=0}^{\infty} = \frac{-1}{\beta(u-\frac{1}{\beta})} = \frac{1}{1-\beta u}$$

Expanding in terms of u at 0 yields

$$1 + \beta u + \beta^2 u^2 + ...$$

telling us that the mean is β and the variance is $2\beta^2 - \beta^2 = \beta^2$.

The memory-less property

$$\Pr(X - a > x \mid X > a) = \Pr(X > x)$$

(where x and a are positive) is verified by:

$$\Pr(X-a \quad > \quad x \mid X>a) = \frac{\Pr(X-a>x \cap X>a)}{\Pr(X>a)} =$$
$$\frac{\Pr(X>x+a)}{\Pr(X>a)} = \frac{e^{-(x+a)/\beta}}{e^{-a/\beta}} = e^{-x/\beta} = \Pr(X>x)$$

Gamma:

Since it's defined as a sum of k independent RVs of the exponential type, its mean and variance are $k\beta$ and $k\beta^2$ respectively, and its MGF is $(1-\beta u)^{-k}$. To derive its pdf, we start with $k=2$ and do a convolution of two exponentials, thus:

$$\frac{1}{\beta^2}\int_0^y e^{-x/\beta}e^{-(y-x)/\beta}dx = \frac{ye^{-y/\beta}}{\beta^2}$$

Convolution of this and another exponential ($k=3$ case):

$$\frac{1}{\beta^3}\int_0^y xe^{-x/\beta}e^{-(y-x)/\beta}dx = \frac{y^2e^{-y/\beta}}{2\beta^3}$$

And one more time ($k=4$):

$$\frac{1}{2\beta^4}\int_0^y x^2e^{-x/\beta}e^{-(y-x)/\beta}dx = \frac{y^3e^{-y/\beta}}{3!\beta^4}$$

which makes it obvious that, in general:

$$f(x) = \frac{x^{k-1}e^{-x/\beta}}{(k-1)!\beta^k}$$

for $x>0$ (zero otherwise). To verify, let's find the corresponding MGF:

$$\frac{1}{(k-1)!\beta^k}\int_0^\infty x^{k-1}e^{-x/\beta+xu}dx = \frac{1}{\beta^k(\frac{1}{\beta}-u)^k} = \frac{1}{(1-\beta u)^k}$$

(check!).

The distribution function is

$$F(x) = \Pr(X \le x) = 1 - \left[1 + \tfrac{x}{\beta} + \tfrac{x^2}{2\beta^2} + \tfrac{x^3}{3!\beta^3} + ... + \tfrac{x^{k-1}}{(k-1)!\beta^{k-1}}\right]e^{-x/\beta}$$

To verify this, differentiate $F(x)$ with respect to x, and get $f(x)$.

Multinomial:

It is a generalization of binomial distribution, except: instead of two possible outcomes, each trial can have three (or more - our formulas will assume three). We will call them Win, Loss, and Tie. Then, the probability of i wins, j losses and k ties is computed by

$$\Pr(X=i \cap Y=j \cap Z=k) = \tbinom{n}{i,j,k}p_x^i p_y^j p_z^k$$

can be proven is the same manner as we did for binomial (the sample space would now consist of all n-letter words built out of *three* letters, ...). The marginal distributions of X, Y and Z are (quite obviously) all binomial, so we can easily compute their means and variances. To find $\mathrm{Cov}(X, Y)$, we write

$$\begin{aligned} X &= X_1 + X_2 + ... + X_n \\ Y &= Y_1 + Y_2 + ... + Y_n \end{aligned}$$

where X_1, X_2, ... is the number of wins in Game 1, Game 2, ... (similarly Y_1, Y_2, ... count the losses). Obviously, each of these $2n$ RVs can have only two values, 0 or 1. Now

$$\begin{aligned} &\mathrm{Cov}(X_1 + X_2 + ... + X_n, Y_1 + Y_2 + ... + Y_n) = \\ &\sum_{i,j=1}^{n} \mathrm{Cov}(X_i, Y_j) = \sum_{i=1}^{n} \mathrm{Cov}(X_i, Y_i) + \sum_{i \neq j}^{n} \mathrm{Cov}(X_i, Y_j) = \\ &n \cdot \mathrm{Cov}(X_1, Y_1) \end{aligned}$$

since, when $i \neq j$, the RVs are independent and have 0 covariance. Since $\mathbb{E}(X_1 Y_1) = 0$ (the $X_1 Y_1$ product cannot have any other value than 0, as you cannot have a win and loss in Game 1 at the same time!), $\mathrm{Cov}(X_1, Y_1) = 0 - p_x p_y$. The final formula thus reads:

$$\mathrm{Cov}(X, Y) = -n p_x p_y$$